Three essays on Human Capital: The Role and Determinants of Cognitive and Non-Cognitive Skills

Mario Fiorini

A dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Department of Economics

University College London

April 2008

Declaration

I, Mario Fiorini, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Chapter 3 was undertaken as joint work with Valentino Dardanoni and Antonio Forcina.

ACKNOWLEDGEMENTS

I have a debt of gratitude to my supervisors, Jérôme Adda and Richard Blundell, for their guidance, patience and support, both professional and personal. Jérôme rescued me when in the middle of my PhD I had lost direction. He was willing to supervise me for my job market paper and I have learnt a good deal from him. His door was always open. Richard has been of constant support throughout the whole journey. Indeed, I still remember the day I knocked to his door saying I wanted to do my PhD at UCL and asking him to supervise me. While leaving his office I apologized for my intrusion. He replied "Don't worry, we like enthusiastic people here".

Over the past few years I have also greatly benefitted from conversations with Lars Nesheim, Pedro Carneiro and Valentino Dardanoni. Lars proved invaluable when, with the job market approaching and in Jérôme's absence, he provided me with constant supervision. Without his help my matlab code would still be running in vain. Pedro is one of the humblest academics I have met. He was willing to listen to my ideas for hours, at any time, yet providing great advice. As for Valentino, for several reasons he should have gone top of the list. He is the one to blame if I have become an economist. I met him when I was an undergraduate student in Palermo, and he is one the best teachers I have ever had. But he is also a great researcher and I have greatly enjoyed writing one of the chapters of this thesis with him. Special thanks go to Ian Preston, Wendy Carlin, Stephen Smith, Martin Cripps and Antonio Forcina.

Among my fellow PhD students, a special mention goes to my office mates: Giovanni Gallipoli, Marco Cozzi, Pierre Hoonhout, Giacomo De Giorgi, Marco Angrisani and Esteban Jaimovich. Fantastic friends and fantastic economists. It was fun. And I will always bring with me the moments with Miguel Foguel. I also feel grateful to Albrecht Glitz, Chiara Binelli, and my friends Francesco, Giorgio, Ian, Belen, Noelle, Chiara and Christian in no particular order.

Katrien is the best thing these years in London have given me. She is my star. We shared all we had. And I hope we will keep sharing it for long. Last but surely not least, I am most indebted to my father and my mother who made this possible.

ABSTRACT

This dissertation analyzes the role and determinants of cognitive and non-cognitive skills. A number of papers have stressed that educational and labor market outcomes are largely predetermined by the cognitive and non-cognitive skills accumulated during early childhood. Some of these papers recommend investing in this type of skills to raise educational enrolment and attainment, to reduce disparities between ethnic groups or to weaken the intergenerational transmission of socio-economic status. Yet a number of questions are still open: Is early investment in skills always the best option? Do cognitive and non-cognitive skills account for most of the intergenerational transmission of socio-economic status? What are the most important inputs of these skills? The first essay compares the efficiency of two alternative policies aimed at fostering educational enrolment. The results indicate that a direct grant in the form of a tuition subsidy is more efficient than an equally expensive unconditional parental income subsidy given when individuals are still in their childhood. The shift in the cognitive skills distribution following the latter subsidy is too small to generate a large increase in college enrolment. The second essay tests for stochastic monotonicity in intergenerational socio-economic mobility tables. The results provide evidence of monotonicity both unconditional and conditional on educational attainment, cognitive and non-cognitive skills. The third essay shifts the attention to the determinants of these skills, and in particular to the effect of using a computer at home on children's development. The results indicate that time spent on the computer has a positive effect on cognitive skills. For the non-cognitive skills the evidence is more mixed, with the direction of the effect depending on the type of skill and the age of the children.

CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

In the last decade a number of papers have stressed that educational and labor market outcomes are largely pre-determined by the cognitive and non-cognitive skills accumulated during early childhood. Keane and Wolpin (1997, 2001) and Cameron and Heckman (1998, 2001) have found that in the US individual educational decisions are mainly driven by cognitive skills such as maths and verbal skills. Those with good skill endowments by age 16 are much more likely to enrol and complete college education. Financial constraints are either not binding or most individuals manage to offset them by working part-time and borrowing. Their results suggest that policies targeting educational attainment or educational disparities between Black, Hispanic and White youth must act on these skill inputs to be effective. Heckman, Stixrud, and Urzua (2006) find that a low-dimensional model of cognitive and non-cognitive abilities explains a diverse array of outcomes such as schooling choices, wages, employment, work experience, choice of occupation but also a variety of adolescent risky behaviors such as criminality, cigarette smoking and alcohol use.

Chapter 2 investigates how educational choices are affected by cognitive skills accumulated during childhood. In particular, a specific policy question is addressed *"If the Government aims to increase enrolment in post-compulsory education which policy is more effective? should the Government subsidize parents at an early stage, to increase their investment in the child's cognitive skills during childhood, or should it instead subsidize individuals directly through grants in post-compulsory education?"*. Previous studies have tried to evaluate the effectiveness of financial incentives and short-term liquidity constraints at age 16 or later, but rarely found them to be relevant for educational and occupational choices. Their concluding remarks pointed to ex-ante heterogeneity as the main determinant of these choices. However not many papers have compared the effectiveness of alternative policies. The main contribution of this chapter is to test the effect of early policy intervention on cognitive skills and evaluate the impact on educational decisions. To tackle this issue I exploit the information provided by the National Child Development Survey (NCDS), a U.K. cohort study following individuals from birth onwards that collected information on skills and parental background information at different stages of childhood. When targeting an increase in Higher Education enrolment equal to 1% of the population, the results indicate that a subsidy (grant) at the age of 18 is the most efficient way. A subsidy of the same amount, but given to parents when the individual is still in its childhood, would instead increase cognitive skills and in turn enrolment into Higher Education. However enrolment would not increase as much as after a direct Higher Education subsidy at the age of 18. This result does not imply that additional investment in cognitive skill accumulation is wrong, but that such investments should be well structured and ensure a high return, otherwise they risk being inefficient.

If childhood is so important for later labor market outcomes than surely parental investment plays a key role. Starting with Becker and Tomes (1979) researchers have proposed economic models of intergenerational mobility to uncover the mechanism behind the transmission of social status. Becker and Tomes (1986), Solon (1999), Mulligan (1999), Han and Mulligan (2001) and Restuccia and Urrutia (2004) are all attempts in that direction. Currently, it is widely accepted that parental transmission of skills, beliefs, motivation and social connections are all important in explaining the strong dependence between the social status of father and son.

Chapter 3 is joint work with Valentino Dardanoni and Antonio Forcina. Our objective is to test for stochastic monotonicity in intergenerational socio-economic mobility tables, that is whether having a parent with a higher socio-economic status is never worse than having one with a lower status. We first apply the test to a set of 149 intergenerational mobility tables in 35 different countries, where it emerges that monotonicity cannot be rejected in almost any table. We then address the following question: *Does stochastic monotonicity still hold once we account for education, cognitive and non-cognitive skills?* One main contribution of this chapter is to formalize and apply tests of dependence in mobility tables, using both continuous and discrete control variables. Our results using the UK NCDS cohort data confirm that education, cognitive and non-cognitive skills do explain a substantial share of the mobility mechanism. Nevertheless we find evidence of monotonicity both unconditionally and conditionally.

Given this evidence on the importance of cognitive and non-cognitive skills, there is a growing interest in estimating the skills production function. Researchers are trying to uncover the main inputs and their time varying effect (see Todd and Wolpin (2003, 2007) and Cunha and Heckman (2007) for a discussion). However estimating the causal effect of these inputs is difficult because all sorts of endogeneity problems might lead to inconsistent estimates and economists have mainly focused on a few inputs that are either very important or for which experimental designs are available. To mention only a few recent studies that have looked at the determinants of math and reading achievements, Rivkin, Hanushek, and Kain (2005) analyze the effect of teacher quality, Dahl and Lochner (2005) and Belley and Lochner (2007) estimate the effect of parental income, Bernal and Keane (2008) and Berlinski, Galiani, and Manacorda (2008) evaluate the effect of respectively child care and pre-primary education while Gentzkow and Shapiro (2008) identify the effect of pre-school television exposure.

Chapter 4 investigates the effect of using a computer at home on children's development. In most OECD countries 70% or more of the households have a computer at home and children use computers quite extensively, even at very young ages. Yet, little is known about the effect of computer usage on children's cognitive and non-cognitive skills. This chapter seeks to answer the next question: *Is time spent using a computer an important input of the skills production function?* Time spent using a computer can affect skills because of the way children use the computer, i.e. content, because computer time inevitably displaces other activities, and because most software requires interaction and is therefore intellectually stimulating. I use data from the Longitudinal Study of Australian Children (LSAC), which follows an Australian cohort born in 2000. Skills and computer usage information is collected when children are approximately 5 and 7 years old. For cognitive skills, the results indicate that computer time has a positive effect. The effect is large relatively to other inputs such as child care, and is not shared by other media devices, such as television and video games which instead produce a negative effect. For the non-cognitive skills the evidence is more mixed, with the direction of the effect depending on the specific skill and the age of the children. I test the robustness of the results comparing OLS, IV and Value Added estimators. Generally, the IV estimates are larger and the Value Added estimates lower than the OLS ones. However the pattern of the results is quite consistent.

CHAPTER 2

FOSTERING EDUCATIONAL ENROLMENT THROUGH SUBSIDIES: THE ISSUE OF TIMING

2.1 Introduction

The objective of this paper is to investigate how educational choices are affected by cognitive skills accumulated during childhood. In particular, we address a specific policy question *"If the Government aims to increase enrolment in post-compulsory education which policy is more effective? should the Government subsidize parents at an early stage, to increase their investment in the child's cognitive skills during childhood, or should it instead subsidize individuals directly through grants in post-compulsory education? "*.

The relevance of this question is proved by recent policies introduced in the United States and in the United Kingdom. Specifically, in September 2004 the U.K. has introduced the Education Maintenance Allowance (EMA) to increase participation in education at the age of 16, the minimum school leaving age. In 2006 the University tuition fee system has changed, the so called Top-Up fees project, with a combination of higher tuition, loans and grants to induce enrolment in Higher Education.

Previous studies have tried to evaluate the effectiveness of financial incentives and short-term liquidity constraints at age 16 or later, but rarely found them to be relevant for educational and occupational choices. Their concluding remarks pointed to ex-ante heterogeneity as the main determinant of these choices. This ex-ante heterogeneity has always been treated as exogenous and called skill endowment. [1]

[1] In this paper we distinguish between ability, that we consider innate, and skills, that are instead the result

Our contribution is to test the effect of early policy intervention on cognitive skills and evaluate the impact on educational decisions. To tackle this issue we exploit the information provided by the National Child Development Survey (NCDS), a U.K. cohort study following individuals from birth onwards that collected information on skills and parental background information at different stages of childhood. We are therefore able to observe the skills accumulation process together with educational and occupational choices up to the age of 41. These data are unique, because they follow individuals from birth while the majority of other cohort studies, such as the National Longitudinal Survey of Youth (NLSY), follow individuals from the age of 14 or later, therefore lacking childhood information. The NCDS data give us the unique opportunity to test the effect of variation in cognitive skills measured as early as age 7. [2]

In our approach we consider both the investment of parents into the child's cognitive skills, through a production function capturing the causal effect of parental income on skills, and the educational choices determined by preferences and accumulated skills.

We model the educational choices with a dynamic structural model. The advantage of this approach lies in the possibility of inferring the preferences and beliefs of individuals, modelling the selection into schooling based on observables and unobservables and leaving us the option of evaluating the effect on educational decisions of changes in key parameters such as the distribution of cognitive skills and tuition fees.

Since we do not have a plausible exclusion restriction, identifying the return to parental income in cognitive skills is problematic. What we do instead is to use the structure of the model and the comparison of the two education subsidies, the tuition subsidy (grant) and the parental income subsidy, to identify which return to parental income would make the two subsidies equivalent. [3] We focus on men as the educational decisions and wage outcomes are gender specific. We further restrict the sample by excluding self-employed individuals since their wages are not always well reported.

Finally, our set up is a partial equilibrium model, and therefore does not take into account

of ability and the learning process. Therefore ability and skills are identical only at birth.

[2]The NCDS follows individuals born in 1958 matching closely the NLSY79 which follows individuals born between 1957 and 1965.

[3]The Data include some measure of investment such as parental interest in the child's education (assessed by school teachers) but, the aim of the paper being to simulate alternative policies, it is not straightforward to imagine how this form of parental investment could be affected by Government policy. What we have in mind are financial forms of parental investments that could be changed through monetary subsidies.

the consequences of increased skill levels or higher qualifications on their relative returns. As an implication our model can be used to assess the effect of only relatively small changes in the stock of human capital.

When we target an increase in Higher Education (college or equivalent) enrolment equal to 1% of the population, our results indicate that a subsidy (grant) at the age of 18 is the most efficient way. The same subsidy, given to parents when the individual is still in its childhood, would increase cognitive skills and in turn enrolment into Higher Education, but not as much as a direct Higher Education subsidy at the age of 18. The same conclusion is reached if we instead target an increase in education enrolment between age 16 and 18.

Our result does not imply that additional investment in cognitive skill accumulation is wrong, but that such investments should be well structured and ensure a high return. Otherwise they risk being inefficient.

The paper is organized as follows: In section 2.2 we review the main findings of the literature. The literature on the economics of education being quite vast, we focus on those papers investigating the importance of short term financial constraints and of comparative advantage, which are the most relevant to our policy question. Section 2.3 describes the UK education system. Section 2.4 presents the cohort data that we use. In Section 2.5 we go through the economics of the model, while in Section 2.6 and Section 2.7 we explain respectively its identification and estimation. Section 4.5 first presents the maximum likelihood estimates and the model fit, and then it describes our simulations. Section 4.6 concludes.

2.2 Related Literature

The recent US literature on educational choices has looked at the importance of financial incentives and budget constraints in schooling decisions.

Keane and Wolpin (1997) investigate the educational and occupational decisions of a US male cohort born in the 60's (NLSY 79) in a dynamic discrete choice structural model estimating the impact of a tuition fees subsidy on the college participation decision. In the model individuals are not financially constrained and from age 16 onwards decide year by year whether to stay in education, stay home, work in a white or blue collar occupation. The model allows for ex-ante (age 16) heterogeneity in endowment. Keane and Wolpin find that a $2000 college fees subsidy would increase high school graduation by 3.5 percentage points and increase college graduation rates by 8.4 percentage points. However, when analyzing the life-time utility effects, those who would benefit most from the subsidy are the individuals with high endowment for school education and white collar occupations, who would have gone to college even without the subsidy. Those induced to go to college by the subsidy are individuals with low endowment for education and a comparative advantage in blue collar occupations, with only a minor increase in their lifetime utility. This is due to part of the subsidy being spent to compensate the pre-policy larger utility from a no-college choice. They conclude that ex-ante heterogeneity in skill endowment is a major determinant of responses to subsidies and effects in lifetime utility.

Keane and Wolpin (2001) extend their previous work to account for financial constraints at the age of 18, with the model also allowing for parental transfers from age 16 onwards and marriage as important factors affecting the decisions. They find that borrowing constraints exist and are tight but have a limited effect on college attendance decisions, since individuals adjust their behavior working part-time or reducing consumption while at school. Hence, subsidizing poor parents would have little effect on college participation. The latter results support the hypothesis that ex-ante heterogeneity plays a central role in educational and occupational decisions.

Cameron and Heckman (1998) use an ordered choice dynamic model to explore whether the importance of family background factors in educational decisions decreases as individuals move towards higher grades. They look at five US cohorts born between 1907 and 1964. They find that these background factors have a rather constant importance across cohorts, but that family income is not so relevant once controlling for observed cognitive skills. Their conclusion is that parental factors are important mainly because they shape child's skills and taste for education early in life, and these latter characteristics determine educational choices.

Cameron and Heckman (2001) estimate a dynamic model of schooling attainment in the US, investigating the sources of educational disparities between Black, Hispanic and White Males. While it is often found that these disparities are linked to parental income differentials, the paper tests whether this effect is due to long-term effects or short term financial constraints. They estimate the model separately for the 3 ethnic groups and find that parental income is important, but its effect is largely diminished once they control for AFQT scores. They also test the effect of variation in costs, equalizing the tuition fees across ethnic groups, but again do not find large effects. On the other end, equalizing the AFQT, would lead to blacks and hispanics having higher enrolment rates than whites.

Carneiro and Heckman (2002) look at the same US cohort as the Keane and Wolpin studies, and investigate in a static model the importance of short run and long run factors influencing the college attendance decision. Short run factors are associated with liquidity constraints at age 18, while long run factors are linked to permanent differences due to parental background. Once again, conditioning on skills measured in early teenage years, short term constraints play only a minor role. Their results suggest that at most 8% of American youth face short term liquidity constraints that affect post secondary schooling.

Cameron and Taber (2004) measure the importance of borrowing constraints on education decisions. Their intuition is that opportunity costs and direct costs of schooling affect borrowing constrained and unconstrained persons differently. Direct costs need to be financed during school and impose a large burden on credit-constrained students. By contrast, gross forgone earnings do not have to be financed. They explore this idea using both a reduced form IV strategy and a structural model approach. However, in no case they find evidence of borrowing constraints.

Dearden, McGranahan, and Sianesi (2004) replicate the analysis of Carneiro and Heckman (2002) using UK data, the National Child Development Survey (NCDS) data and find that once controlling for test scores in mathematics and reading, individuals do not seem to suffer from short-term credit constraints. Their findings suggest that policies aimed at reducing the impact of credit constraints on education decisions should target individuals at the age of 16 (or possibly earlier) when staying-on decisions are made, rather than at age 18 when individuals are making Higher Education decisions. [4] However, given the reduced form approach and no information on schooling direct costs, they can not simulate the possible impact of a Government financial subsidy to parents or individuals on their educational decisions.

Attanasio, Fitzsimons, and Meghir (2004) apply a dynamic structural model with ex-ante (age 16) heterogenous individuals to estimate the impact of the Education Maintenance Allowance on

[4] An example of this policy is the Education Maintenance Allowances programme.

the decision to either stay in education, stay in education and work part time, or leave education and work full time. They use the data provided by the experiment conducted in 1999 when the EMA was introduced in ten Local Education Authorities. This survey followed individuals for only 3 consecutive years from age 16. They find that the EMA program increased participation in education without a part time job from age 16 to 18 and increasing the generosity of the EMA would further augment such participation. However, the EMA would have only negligible effects on the participation in Higher Education. [5]

All together these studies provide a strong indication that policies aimed at increasing education attendance through monetary subsidies might not be very effective. Only a small fraction of individuals appears not to enroll because of short term financial constrains, and those who enroll because of the subsidy do not have big gains in utility. Instead, these papers point to the comparative advantage hypothesis as originally described by Roy (1951) and subsequently by Willis and Rosen (1979). If an individual has accumulated enough skills then staying in education will be rewarding, while if the individual did not accumulate the skills staying on would actually lead to lower utility than otherwise obtained entering the labor market immediately.

An education subsidy could change the participation decision for those at the margin, but part of the subsidy would be lost in compensating the difference in utility caused by the comparative advantage.

The policy recommendation for a Government aiming to increase enrolment into education is to intervene not at age 16-18, when staying-on decisions are made, but during childhood, when individuals accumulate their skills, cognitive and non-cognitive. The same subsidy given at the age of say 11 could foster the accumulation of skills and be therefore more effective on schooling choices then if given at age 18. Nevertheless, these studies do not compare the effectiveness of the two alternative policies: the fees subsidy (grant) and alternative early intervention.

Estimating the impact of Government intervention so early in life is, however, a hard task given the shortage of surveys monitoring parental decisions and skills accumulation. The NLSY longitudinal data used in many of the papers mentioned above follow individuals at best from the age of 14 onwards, with a measure of skills given by the AFQT test score and with parental

[5] The EMA was only given until age 18.

income measured at the age of 16. Therefore it does not allow testing the impact of subsidy to individuals or their parents anytime before.

2.3 The UK Education System

Before introducing the model, we give a brief overview of the UK Education System in order to understand the individual decision process and the assumptions behind our model.

We focus on the English and Welsh education system, which are identical, and omit the Scottish Education System, which is slightly different from the previous ones. This choice is driven by the difficulty of modelling the educational decisions for both types of systems.

Schooling is compulsory up to the age of 16, when individuals can, at the end of the scholastic year, stay in education or enter the labor market. If they stay, there are two main educational paths that they can follow: the Academic and the Vocational one. We describe both in turn although they are not necessarily mutually exclusive, i.e. individuals might take both Academic and Vocational Qualifications.

We discuss the system faced by individuals born in 1958 which, with some differences, also reflects the current one.

2.3.1 The Academic Path

The Academic path is mainly full-time. Those who stay on at age 16 enrol for the O Levels or CSE qualifications, which are taken immediately at the end of the scholastic year. These students are still aged 16 when they obtain the qualification. O Levels are single subject examinations reflecting the single disciplines of the university departments and faculties. They are designed for more able secondary school students and are necessary for progression into further education (A-level). The Certificate of Secondary Education (CSE) qualification was intended for students whose skills were not considered sufficient for O level courses. Nevertheless, there was an overlap between these two types of certificates in that a CSE grade 1 result was regarded as equivalent to an O-level pass. Even though there was no formal requirement, students would be expected to pass at least 5 O Levels graded A-C or CSE graded 1st, in order to stay in education afterwards.

In the Autumn term of the same year, those who successfully obtained 5 or more O Levels/CSE can enrol for A Levels. These would last 2 years, until individuals are aged 18. Advanced levels have their origins in qualifications constructed by groups of universities in the first half of the century, designed to identify candidates suitable for degree courses and to provide a foundation for advanced teaching in a single subject area to degree level. These characteristics still distinguish A levels. They are still awarded by independent examining bodies with close links to the universities. Passing 2 A-level constitutes the minimum level required for entry in Higher Education. Normally two or three A-levels are studied.

Once the student has completed A Level, he can gain admission to the Universities, Polytechnic or Colleges of Higher Education where a first degree is obtained. The time needed to gain a degree varies by subject but in the majority of cases it takes 3 years.

Therefore a student who completes the Academic path with no interruption will normally enter the labor market at the age of 21 or 22. [6] Hereafter we use OL, AL and HE for O Levels, A Levels and Higher Education respectively. [7]

2.3.2 The Vocational Path

The Vocational Path is quite different from the Academic one, mainly in the types of qualification awarded and in the timing. Vocational qualification are more specialistic and often linked to the acquisition of a competence requested for a specific kind of job. They range from advanced food technology, catering degrees to lower levels dog grooming and cake decoration ones. Some of these qualification are taken while in full-time education while others are taken later on in life, even after entering the labor market. Yet, vocational qualifications are grouped in O Level equivalents, A Level equivalents and Higher Education equivalents to match them with the Academic ones.

[6] Individuals might stay in education even further to complete a post-graduate degree. However this was very unlikely among the older cohorts such as the NCDS one.

[7] Another important feature of the British education system was the early selection of pupils into different types of schools. Prior to, and during the 1960s, the British education system was selective. Pupils were tracked into different schools, according to their ability at age 11. The most able, who passed an entrance examination at age 11, went to grammar schools. The rest went to secondary modern or technical schools. The selection of around only 20 per cent of the cohort for a grammar school education (which could lead to university entrance) was progressively challenged in the 1960s. From 1965 onwards (Circular 10/65), Labour governments encouraged local authorities to develop comprehensive schools which accepted all children from a neighborhood, regardless of ability. Comprehensive education slowly gained ground and by the end of the 1970s over 80 per cent of all children in maintained schools were in comprehensive schools.

2.4 Data and descriptive statistics

The National Child Development Study (NCDS) targets over 17,000 babies born in Britain in the week 3-9 March 1958. Surviving members of this birth cohort have been surveyed on six further occasions in order to monitor their changing health, education, social and economic circumstances: in 1965 (age 7), 1969 (age 11), 1974 (age 16), 1981 (age 23), 1991 (age 33) and 1999 (age 41). At the age of 7, 11 and 16 mathematics, reading and general skills tests were taken by the cohort member. Information about parental background including education and occupation was also collected during those years, with a measure of parental income included at the age of 16. Moreover, in 1978, when individuals where aged 20, a survey was conducted among the secondary schools where they had taken their qualifications up to A Levels. This school survey allows to distinguish between enrolment in OL and AL and the actual achievement of the qualification.

Although the surveys were not conducted on yearly basis, information on labor market history, including employment status and occupation, was gathered for each month of the cohort members life from age 16 onwards through retrospective questions. Data on wages instead were only collected at the time of the interviews.

These data sets therefore bring together information on educational and occupational choices, skills and parental background measures collected in the childhood years of the cohort members.

We select all males for which we observe parental income, skills test scores at age 16 and educational choices.

Table 2.1 shows the coding and the composition of our sample by status and age. We select those individuals observed for at least two periods. Given the education system we start our analysis with the period April 1974 up to September 1974, entry 16^a in our table, when individuals were already 16 and could leave education. This is the moment we can observe their first decision. Because this period is mainly within the scholastic year and summer holidays, we assume that individuals either enrolled in education or stayed home, and are therefore assumed to be unemployed. The following period runs from October 1974 until September 1975, entry 16^b, when individuals could have studied A Levels, found a job or be unemployed. We classify individuals as employed if the number of months spent working was larger than time in unemployment. This period and all the following ones are yearly time periods, starting in October and ending in September of the following year. Above age 21, that is after O Levels, 2 years of A Levels and 3 years of Higher Education individuals are supposed to be in the labor market. [8] We do not model re-entry in education. Among those in full-time education more than 93% of the individuals did not have any break from full-time education, where a break is arbitrarily defined as a six months or longer period outside full-time education before going back. This percentage raises to 95% when we consider a break as 12 months or longer period. We drop those individuals who have a break of 12 months or longer.

From the data it is clear that the unemployment rate in the sample rose suddenly around the age of 22, from 5% to 10%, only to go down again two years later, at age 24. The work rate mirrored the unemployment one given that all individuals were already in education. This was the time when Margaret Thatcher was prime minister with improved productivity but soaring unemployment in the UK. What casts a shadow though is the peak in the series. Since individuals were interviewed at the age of 23 and 33, the reported unemployment from the age of 24 onwards comes from the age 33 survey, with a risk that individuals under-reported their unemployment in those years, particularly if that occurred for short periods. [9]

Figure 2.1 shows the educational choices and the obtained qualifications. Although 63% of men in the sample enrolled for O Levels, only 28% obtained sufficient grades to progress to A Levels. Part of these individuals left education anyway. In the end only 7.3% of the sample obtained a HE qualification. Note that we do not distinguish between individuals who enrolled and individuals who successfully completed Higher Education. This is because, while data were collected from secondary schools, no data were collected from HE institutions. In our sample therefore we code as being enrolled all those individuals that reported having a HE qualification. [10]

[8] By the age of 23, less then 1% reported having a post graduate qualification. This percentage raised to almost 4% by the age of 33. We exclude these individuals from our sample.

[9] We have done some investigation on this, though we could not find a reliable unemployment statistic in the early 80's for those born in 1958. (i.e. a cohort specific statistic.) Nevertheless using available statistics or some other data set such as the FES, it seems that my data underestimate unemployment by around 3% between the age of 24 to 26.

[10] We have also tried to investigate this issue using some self-reported data at age 23, but the results are not fully convincing. It seemed that up to 16% of individuals might have failed an Higher Education course. It is hard to say whether we could do more on this because not all individuals in the sample actually took part in the age 23 survey, and for the missing ones we would not know whether they failed an HE course.

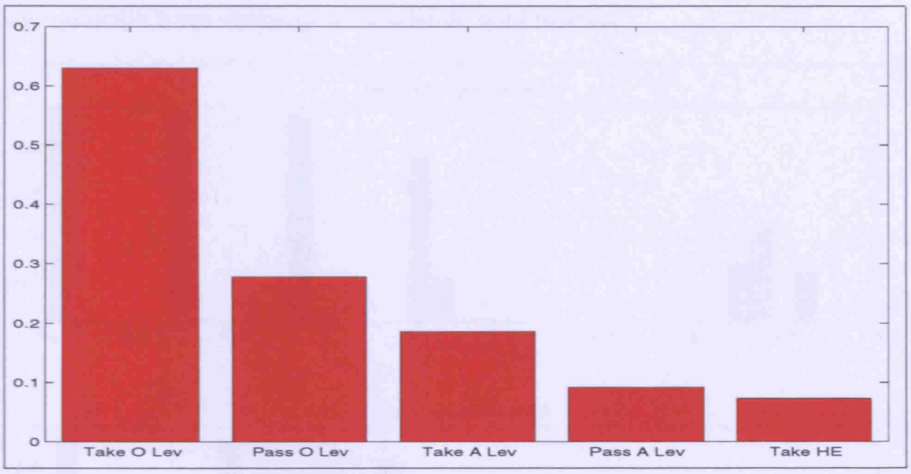

Figure 2.1: Educational choices

Table 2.3 shows sample statistics for the mathematics and reading test scores at the age of 16 and 7, and for parental income at the age of 16. At age 7 individuals were administered four tests (reading, mathematics, copying and drawing) while only reading and mathematics tests were administered at age 16. Between age 7 and 16 the tests were changed to take into account the age difference. We only use the mathematics and reading scores. [11] In order to reduce the number of state variables we summarize the mathematics and reading test scores using the first principal component of the standardized test scores. The assumption that one factor captures cognitive skills is quite widespread in the literature and in line with the g theory used by Herrnstein and Murray (1994). Not reported here, the first principal component for the age 16 scores explains 82% of the total variance while the age 7 principal component explains 77 % of the total variance. The loading factors of the standardizes scores are identical and equal to 0.70711, both at age 16 and 7. Higher values of the principal component correspond to higher scores. [12]

Figure 2.2 shows the obtained qualifications by principal component (quintiles). It is clear that both age 16 and age 7 cognitive skills are good predictors of educational achievement. Those

[11] The questionnaires, including the tests can be downloaded from the UK data archive website. The age 7 questionnaire can be found at http://www.data-archive.ac.uk/doc/3148/mrdoc/pdf/a3148uab.pdf, while the age 16 questionnaire can be found at http://www.data-archive.ac.uk/doc/3148/mrdoc/pdf/a3148ucb.pdf.

[12] The loading coefficient of the original, non-standardized scores were also very close and equal to 0.71 and 0.69 respectively for the math and reading scores.

in the lower quintiles have virtually no academic qualification.

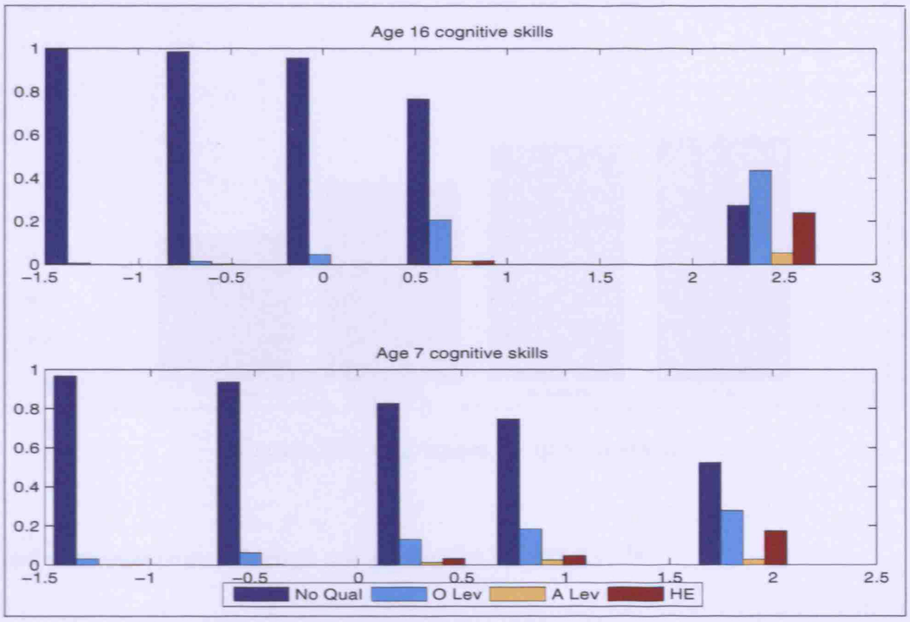

Figure 2.2: Qualifications by Skills.

Figure 2.3 reports mean log yearly wages by obtained qualification (in January 2001 prices). The wages are monotonically increasing. In table 2.2 we report separate statistics for the three age points when we can actually measure wages. In the first line we report the mean wage by age; as expected it is increasing as the individuals grow older. The following lines decompose the first one by highest qualification obtained. Higher qualifications are usually associated with higher wages.

The British NCDS cohort is comparable to the US National Longitudinal Survey of Youth (NLSY79) cohort, composed of individuals born between 1957 and 1965. The NCDS though followed the individuals from birth, while the NLSY first surveyed individuals in 1979 when they were between 14 and 22 years old. The NCDS study reports reading and mathematics test scores at the age of 7, 11 and 16. This is unique. The NLSY79 reports the AFQT test but this was administered in 1980 to all. Individuals in the NLSY were therefore aged between 15 and 23 when they took the test. The NCDS has also age 0, 7 and 11 information on parental

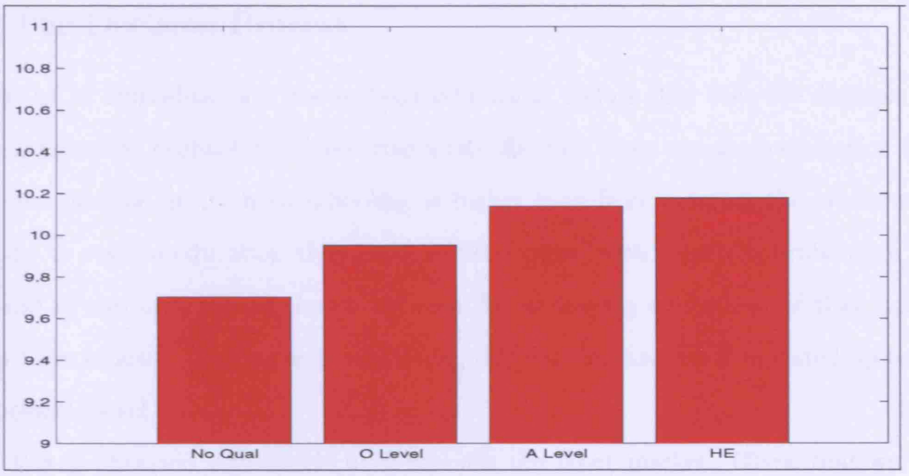

Figure 2.3: Log wages by qualification

background characteristics which is not present in the NLSY. [13]

2.5 *The Model*

In modelling the educational choices we only look at the Academic Path. This is because the qualifications are quite homogenous and the timing is very similar across individuals. In the data only 2% of individuals had Vocational qualifications without any Academic one at the age of 23. Conlon (2001) found that the wage returns to Academic qualifications are usually much larger than to Vocational qualifications. We choose to model an optimal stopping problem, where at the age of 16 (OL and AL) and 18 (HE) individuals decide between staying on or leaving education and enter the labor market. The labor market is an absorbing state. Given the low number of individuals that re-enter education after a year long break, an optimal stopping model should match fairly well the choices made in the Academic path.

[13] The Armed Forces Qualifications Test score (AFQT) is a composite score derived from select sections of the Armed Services Vocational Aptitude Battery (ASVAB), a battery of 10 tests that measure knowledge and skill in the following areas: (1) general science, (2) arithmetic reasoning, (3) word knowledge, (4) paragraph comprehension, (5) numerical operations, (6) coding speed, (7) auto and shop information, (8) mathematics knowledge, (9) mechanical comprehension and (10) electronics information.

2.5.1 The Decision Process

At the age of 16 individuals are free to leave education. Before they take the decision they are assumed to know the probability of receiving a job offer (δ). They remain in education whenever the expected lifetime utility from schooling is higher than from entering the labor market. If they decide to stay in education they enrol in OL exams, with a certain probability (λ) they succeed and in the next period choose between AL or leaving education. If they fail ($1 - \lambda$) then they have to leave. As long as in education, this decision process is repeated up to HE, the highest possible level.

Once this is obtained individuals have to enter the labor market. Given that we focus on men, we assume that once out of education individuals always choose to work. Nevertheless, they could be unemployed if they do not receive a job offer (provided they were in education or unemployed in period $t-1$) or if they get fired (provided they were working in period $t-1$) which occurs with a probability (ϕ). The only decision in the model is between staying or leaving education. The work/unemployement status is not a decision.

This process being sequential, we model it as a discrete Markov decision process (DMDP) where at each point the choice depends only on the current level of the individual characteristics or state space. See Eckstein and Wolpin (1989) and Rust (1994) for a review of DMDP models and estimation strategies.

2.5.2 Instantaneous Utility

We follow Keane and Wolpin (1997) and model the rewards in monetary terms. This choice is dictated by the absence of variation in tuition fees in the data. [14] The linear utility in income implies that the individuals are not liquidity constrained and are risk neutral. In the conclusions we discuss possible implications of these assumptions. Individual and time specific subscripts are suppressed for clarity of exposition.

The reward from work is given by the annual wage:

[14] Ichimura and Taber (2002) discuss semiparametric identification of tuition subsidy effects whenever the researcher can observe variation in tuition. While in England and Wales education at every stage was free, individuals enrolled in Higher Education were entitled to a grant linked to parental income. The amount of the grant was between 200 and 5000 pounds in 2001 prices. Since a tuition subsidy is equivalent to a grant, we could try to use variation across individuals. However, given its link to parental income, it is not easy to identify the effect of the grant net of income.

$$R_W = w = exp\left\{\sum_{j=1}^{3}\alpha_{qj}Q_j + \alpha_x X + \sum_{j=1}^{3}\alpha_{xj}XQ_j + \alpha_{k2}K_{16} + \alpha_{k1}K_7 + \alpha_z + u\right\} \tag{2.1}$$

The effect of schooling is represented by the 3 dummies (Q_j), corresponding to obtained OL, AL and HE qualifications and their interaction with experience (X). The equation also includes skills at age 16 (K_{16}), age 7 (K_7). We interact qualification with experience to better fit the wage profile. We include K_7 since it is a good proxy for innate ability and in the policy counterfactual we test the effect of a change in its distribution.

The unobserved component includes a type specific constant α_z and measurement error u: $u \sim N(0, \sigma_u)$. We model the unobserved heterogeneity with a mixture model and Z types. [15] Provided $\{\alpha_{k2}, \alpha_{k1}\}$ are positive and the return to schooling are monotonically increasing in Q, the exponential functional form ensures a positive cross-derivative between schooling and cognitive skills.

The reward from unemployment is

$$R_U = \overline{R} \tag{2.2}$$

that is individuals receive a fixed amount in unemployment benefit if they are unemployed. $\overline{R}$ is set equal to 3000 pounds, approximately the 1975 benefit in 2001 prices. Until 1995 the benefit was not a function of income. [16]

The latent reward from schooling is given by

$$Rs_j = \gamma_j + \gamma_z + \epsilon_j \quad j = \text{OL, AL, H.E.} \tag{2.3}$$

which indicates that the reward is qualification specific. In the case of O Levels, the first schooling stage, γ_{OL} is set equal to zero because we already have the type specific constant term. The reward is expressed in monetary terms and includes the costs as well as the consumption

[15] See Everitt and Hand (1981), McLachlan and Peel (2000) and Heckman and Singer (1984) for a discussion of mixture models.

[16] Source: Institute for Fiscal Studies , http://www.ifs.org.uk/ff/indexben.php

value of education as measured by the constant term γ_{j0} and γ_z, the persistent unobserved heterogeneity common to all schooling levels. The taste shock (ϵ) is assumed to be i.i.d. and normally distributed $N(0, \sigma_\epsilon)$. Tuition fees are imposed on the model but set equal to zero to reflect free education in the UK at that time. Note that there is no observed heterogeneity in the latent reward from schooling. Cognitive skills are excluded to help identification as we explain later on. This restriction holds if parental transfers and effort are not skill specific, or if they offset each other, i.e. less skilled individuals make more effort but also receive more resources from parents to compensate for their skill gap. We do not include parental income either. Carneiro and Heckman (2002) and Dearden et al. (2004) have found that parental income does not affect educational decisions once cognitive skills are controlled for. Moreover, it is a continuous variable and would increase computational time. [17]

Finally define the latent index generating the probability of obtaining a qualification once enrolled λ, the probability of receiving a job offer δ, the probability of being fired ϕ as

$$\begin{aligned} \lambda_j^* &= \lambda_j + \lambda_{k2} K_{16} + \lambda_{k1} K_7 + \lambda_z + \varrho_\lambda \\ \delta^* &= \sum_{j=1}^{3} \delta_{qj} Q_j + \delta_{k2} K_{16} + \delta_{k1} K_7 + \delta_x X + \delta_z + \varrho_\delta \\ \phi^* &= \sum_{j=1}^{3} \phi_{qj} Q_j + \phi_{k2} K_{16} + \phi_{k1} K_7 + \phi_x X + \phi_z + \varrho_\phi \end{aligned} \tag{2.4}$$

such that the unobserved heterogeneity enters each probability and ϱ is normal.

2.5.3 Solving the Sequential Decision Problem

Define the state space $\Omega = [K_{16}, K_7, Q, X, \epsilon, \alpha_z, \gamma_z, \lambda_z, \delta_z, \phi_z]$ where $[K_{16}, K_7, \epsilon, \alpha_z, \gamma_z, \lambda_z, \delta_z, \phi_z]$ are exogenous states in the model. Define the decision space $D = [Schooling, Non-Schooling]$, the status space $Z = [S, W, U]$ which differs from D since work and unemployment are not a choice. Finally define the choice set $C = \{|D(\Omega)|\}$ because individuals have a choice only if they are still in education and hold a qualification lower than HE. Individuals are therefore ex-ante heterogeneous in observed skills K, persistent unobserved heterogeneity $\{\alpha_z, \gamma_z, \lambda_z, \delta_z, \phi_z\}$ and

[17] Even though cognitive skills do not enter directly the utility of schooling they do affect schooling decisions as we explain later.

serially uncorrelated unobserved taste shocks for education ϵ.

We characterize the problem of a finitely lived individual as maximizing the expected present value of lifetime rewards. The value function V of an individual of age a is defined as the solution to Bellman's equation

$$V(\Omega, a) = \max_{d \in C(\Omega)} \left[R(\Omega, d) + \beta \int_{\epsilon'} V(\Omega', a') \right] \tag{2.5}$$

where we integrate over the support of ϵ only since all the other states have deterministic dynamics. Equation (2.5) also highlights the advantage of not modelling the work/no-work choice since this way we only have a univariate integral.

Status Specific Value Functions

Define by V_S the value function from schooling, V_W the value function from work and V_U the value function from being unemployed. Then

$$V_{WU}(\Omega) = \delta(\Omega) V_W(\Omega) + [1 - \delta(\Omega)] V_U(\Omega) \tag{2.6}$$

is the value function if the individual exits education and enters the labor market.

$$V_{SWU}(\Omega) = \int_{\epsilon} max \left\{ V_S(\Omega), V_{WU}(\Omega) \right\} \tag{2.7}$$

is the value function for an individual deciding between staying-on or leaving education.

Having in mind that OL, AL and HE have a different time length and that ϵ is updated every year, the value function from Schooling is therefore defined as:

$$\begin{aligned}
V_{OL}(\Omega) &= R_{OL}(\Omega) + \beta \left\{ \lambda(\Omega) V_{SWU}(\Omega') + [1 - \lambda(\Omega)] V_{WU}(\Omega) \right\} \\
V_{AL}(\Omega) &= R_{AL}(\Omega) + \beta E_{\epsilon}[R_{AL}(\Omega)] + \beta^2 \left\{ \lambda(\Omega) V_{SWU}(\Omega') + [1 - \lambda(\Omega)] V_{WU}(\Omega) \right\} \\
V_{HE}(\Omega) &= R_{HE}(\Omega) + \beta E_{\epsilon}[R_{HE}(\Omega)] + \beta^2 E_{\epsilon}[R_{HE}(\Omega)] + \beta^3 \left\{ \lambda(\Omega) V_{WU}(\Omega') + [1 - \lambda(\Omega)] V_{WU}(\Omega) \right\}
\end{aligned} \tag{2.8}$$

where for clarity we use Ω rather than Ω' if ϵ is the only state to be updated. An individual's value function is specified as the current reward plus future expected utility, which depends on the probability of success in school λ and the probability of receiving a job offer δ. The next period choice between school and labor market (V_{SWU}) vanishes when entering HE.

From our value function specification it can be seen that we update educational qualification to a higher level (Ω') only if the individual was successful once enrolled. Given what we said in section 2.3, this occurs if he obtained 5 or more OL-CSE at the first stage, 3 or more AL at the second stage. The implication is that enrolling for a qualification but achieving anything less than that has no effects on utility.

It might seem incorrect to assign an annual reward for O Levels even though they do not require an additional year of schooling. Nevertheless, we believe that the decision to take O Levels is actually taken around age 15 because it involves specific preparation to the exams. Hence, we still consider R_S as an annual reward, where the year goes from age 15 to 16. We instead set to zero the working reward for not taking O Levels, since the individual would formally still be in education or unemployed for a few months in the summer.

Value function from work:

$$V_W(\Omega) = R_W(\Omega) + \beta\left\{\phi(\Omega')V_U(\Omega') + [1 - \phi(\Omega')]V_W(\Omega')\right\} \tag{2.9}$$

so an individual working has current utility given by the annual wage and future utility from work or unemployment, weighted by ϕ, the probability of being fired. Here Ω' captures the new level of experience.

Value function from unemployment:

$$V_U(\Omega) = R_U(\Omega) + \beta\left\{\delta(\Omega)V_W(\Omega) + [1 - \delta(\Omega)]V_U(\Omega)\right\} \tag{2.10}$$

We do not assume any depreciation of experience so the state space does not change after a period of unemployment.

Note that there is no randomness in either V_W or V_U since both ϕ and δ do not contain any stochastic component and under the assumption of rational expectations the probabilities are always known by the individual.

2.5.4 Economics of the model

Being in school has a direct utility given by R_S, a cost given by the foregone earnings and a return given by the higher wage productivity. Taking O Levels and A Levels is also valuable because it allows access to higher qualifications (option value). Finally, schooling affects the probabilities of being employed in the labor market. Individuals select themselves in education based on their characteristics. Given the exponential form of R_W and conditional on $\nabla_K R_W > 0$, $\nabla_Q R_W > 0$ as it usually found in the literature, $\nabla_{QK} R_W > 0$. Individuals with high values of K have larger foregone earnings but also larger returns to schooling in the future. They will enroll in education as long as these returns are large enough.

The sign of λ, δ and ϕ gradients with respect to Q and K are also very important in driving the educational decisions. A large value of λ induces individuals to stay longer in education. Enroling in education but failing to get the qualification is very costly in the model because there is no change in human capital. The cost is the foregone wage weighted with the probability of finding a job. Thus if $\nabla_K \lambda > 0$ skilled individuals would be more likely to enrol. The effect of δ and ϕ depends on their interaction with the qualifications. If $\nabla_Q \delta > 0$ and $\nabla_Q \phi < 0$ then individuals have incentives to enrol. The sign of the cross-derivatives $\nabla_{QK} \delta$, $\nabla_{QK} \phi$ determines the selection based on observable skills.

Selection on unobserved heterogeneity works in a very similar way. High wage types (α_z) have larger forgone earnings but also larger returns to schooling if $\nabla_Q R_W > 0$. Enrolment in education is also caused by larger utility of schooling (γ_z), success at the exams (λ_z), larger probability of finding a job (δ_z) if $\nabla_Q \delta > 0$ and a lower probability of being fired (ϕ_z) if $\nabla_Q \phi < 0$.

This class of problems can easily be solved by backward induction. Our problem is particularly simple because u is just a measurement error and because decisions are made only at 3 points in life. Therefore the computation of the max in equation (2.5) occurs only at these points.

2.6 Identification

Magnac and Thesmar (2002) show that without unobserved heterogeneity these models are not non-parametrically identified as long as the following structural parameters are not set: the distribution of unobserved shocks, the discount rate, and the current and future preferences in one reference alternative. When unobserved heterogeneity is introduced, non-parametric identification is prohibitive unless very strong restrictions are imposed. We fix the discount factor β to 0.95 and we impose an exponential function for the utility of work in line with Keane and Wolpin (1997). In the utility of schooling we only impose the additivity of the shocks. Normality is imposed on measurement error, the schooling shock and over the transition probabilities λ, δ and ϕ. We make no assumption on the distribution of unobserved heterogeneity. Overall the model is heavily parametric because we are interested in the effect of different schooling levels and in the selection of individuals based on observable skills.

Parameters in R_W, $(\boldsymbol{\alpha})$, are identified from data on wages and the state variables skills, education and experience. The unobserved heterogeneity (α_z) is identified by cross-section variation in wages, conditional on the states, at each of the three wage points available in the data. Because we model selection on unobservables, the return to schooling, to observed skills and to experience are different from OLS ones. No parameter is estimated in R_U. In R_S the parameters (γ) are estimated to match the proportions enrolling for each qualification. The $\boldsymbol{\sigma_\epsilon}$ vector is identified by a model's constraint: the net opportunity cost has a negative effect on the probability of enrolling. [18]

Identifying the scale of the parameters is necessary to test the effect of variation in tuition fees, given that we do not observe this variable nor the utility of schooling. The identification of $\boldsymbol{\sigma_\epsilon}$ is

[18] To see clearly why, let us focus on the HE participation decision and assume for simplicity that the probability of success in school $\lambda = 1$ and the probability of having a job offer $\delta = 1$. Therefore an individual enrols in HE if:

$$R_{HE}(\Omega) + E\left[\sum_{t=1}^{2} \beta^t R_{HE}(\Omega)\right] + V_W(Q = HE, \Omega, a = 21) > R_W(Q = AL, \Omega) + \beta V_W(Q = AL, \Omega', a = 19)$$

Define the net opportunity cost as

$$A(\Omega) = R_W(Q = AL, \Omega) + \beta V_W(Q = AL, \Omega', a = 19) - V_W(Q = HE, \Omega, a = 21)$$

An individual with A Levels enrols in Higher Education with probability

$$1 - \Phi\left(\frac{A(\Omega) - \widetilde{R_{HE}}(\Omega) - E\left[\sum_{t=1}^{2} \beta^t R_{HE}(\Omega)\right]}{\sigma_{\epsilon,HE}}\right) \tag{2.11}$$

where $\widetilde{R_{HE}}$ is the utility of schooling net of the current realization of the shock. Since $A(\Omega)$ is determined by wage data and its coefficient is implicitly normalized to 1 in (2.11) by the model structure, $\sigma_{\epsilon,HE}$ is identified. The same reasoning can be generalized to identify σ_ϵ for the other educational categories.

helped by excluding (K_{16}, K_7) from the utility of schooling. Formally this exclusion restriction is not needed given the non-linearity of V_{WU}, but without a restriction multi-collinearity would lead to large standard errors. The unobserved heterogeneity in the utility of schooling is identified by cross-section variation in schooling choices, conditional on the state variables, at each of the three schooling stages.

The parameters in λ, vector $(\boldsymbol{\lambda})$, are identified merging individual characteristics to the school data, the latter providing information on the number of OL and AL taken and those obtained. Here the unobserved heterogeneity is identified through cross-section variation, conditional on the state variables, in O Levels and A Levels exam success. The parameters in δ, vector $(\boldsymbol{\delta})$, and ϕ, vector $(\boldsymbol{\phi})$, are identified from yearly data on employment status. Unobserved heterogeneity is identified by cross-section variation in the transitions, conditional on the states, at each point in the labor market. Once again the returns to schooling, to skills and to experience take into account selection on unobservables and are therefore different from reduced form estimates.

2.7 Estimation

Define with L_w the likelihood from the wage density, with $L_\lambda, L_\delta, L_\phi$ the likelihoods from the probability of success in school, having a job offer and being fired, and with L_{OL}, L_{AL}, L_{HE} the likelihoods from the probability of enrolling in O Level, A Level and Higher Education. In the absence of unobserved heterogeneity and with uncorrelated error terms the log likelihood would be:

$$\begin{aligned} \ell(\boldsymbol{\Omega_0}, \boldsymbol{\Theta}) &= \ln \prod_{i=1}^{N} (L_w \times L_\lambda \times L_\delta \times L_\phi \times L_{OL} \times L_{AL} \times L_{HE}) \\ &= \sum_{i=1}^{N} (\ell_w + \ell_\lambda + \ell_\delta + \ell_\phi + \ell_{OL} + \ell_{AL} + \ell_{HE}) \end{aligned} \tag{2.12}$$

Given the additivity of $\ell(\boldsymbol{\Omega_0}, \boldsymbol{\Theta})$, estimation could be easily carried out by fast sequential maximum likelihood estimation using the backward induction nature of the problem. The $\{\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\phi}\}$ parameter vectors could be identified by running separate likelihood maximization of L_w, L_λ, L_δ and L_ϕ. Then we could sequentially maximize $L_{HE}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\phi}}, \boldsymbol{\gamma_{HE}})$, solving the problem of an individual with A Levels who chooses whether to enroll in Higher Education, given the known wage returns to schooling and experience, and job probabilities. With γ_{HE} estimated, $L_{AL}(\hat{\alpha}, \hat{\lambda}, \hat{\delta}, \hat{\phi}, \gamma_{\hat{H}E}, \gamma_{AL})$ and then $L_{OL}(\hat{\alpha}, \hat{\lambda}, \hat{\delta}, \hat{\phi}, \gamma_{\hat{H}E}, \gamma_{\hat{A}L}, \gamma_{OL})$ could be maximized sequentially with the same logic. The inconsistent standard errors, due to the estimation error, could be corrected with one Newton step over the whole likelihood. (see Rust (1994) for a discussion on sequential estimation.)

However, when we introduce unobserved heterogeneity the log likelihood becomes:

$$\begin{aligned}\ell(\boldsymbol{\Omega_0}, \boldsymbol{\Theta}, \boldsymbol{\pi}) &= \ln \prod_{i=1}^{N} \left(\sum_{z=1}^{Z} \pi_z (L_w^z \times L_\lambda^z \times L_\delta^z \times L_\phi^z \times L_{OL}^z \times L_{AL}^z \times L_{HE}^z) \right) \\ &= \sum_{i=1}^{N} \ln \left(\sum_{z=1}^{Z} \pi_z (L_w^z \times L_\lambda^z \times L_\delta^z \times L_\phi^z \times L_{OL}^z \times L_{AL}^z \times L_{HE}^z) \right)\end{aligned} \tag{2.13}$$

with π_z being the proportion of individuals of type z. Equation (2.13) can no longer be estimated sequentially, because now we have the log of a sum.

However, Arcidiacono and Jones (2003) show how to extend the Expectation Maximization (EM) algorithm to solve the likelihood maximization sequentially. The EM algorithm consists of two steps:

- E Step. Compute the conditional probability of being the zth type:

$$P(z|\boldsymbol{\Omega_0}, \boldsymbol{\Theta}, \boldsymbol{\pi}) = \frac{\pi_z (L_w^z \times L_\lambda^z \times L_\delta^z \times L_\phi^z \times L_{OL}^z \times L_{AL}^z \times L_{HE}^z)}{\left[\sum_{z=1}^{Z} \pi_z (L_w^z \times L_\lambda^z \times L_\delta^z \times L_\phi^z \times L_{OL}^z \times L_{AL}^z \times L_{HE}^z)\right]} \tag{2.14}$$

- M Step. Estimate $\boldsymbol{\Theta}$ by maximizing the expected likelihood function and update the vector of unconditional probabilities $\boldsymbol{\pi}$, holding the conditional probabilities fixed.

$$\begin{aligned}\boldsymbol{\Theta} &= \arg\max \sum_{i=1}^{N} \sum_{z=1}^{Z} P(z|\boldsymbol{\Omega_0}, \boldsymbol{\Theta}, \boldsymbol{\pi}) \ln \left(L_w^z \times L_\lambda^z \times L_\delta^z \times L_\phi^z \times L_{OL}^z \times L_{AL}^z \times L_{HE}^z \right) \\ &= \arg\max \sum_{i=1}^{N} \sum_{z=1}^{Z} P(z|\boldsymbol{\Omega_0}, \boldsymbol{\Theta}, \boldsymbol{\pi}) \left(\ell_w^z + \ell_\lambda^z + \ell_\delta^z + \ell_\phi^z + \ell_{OL}^z + \ell_{AL}^z + \ell_{HE}^z \right)\end{aligned} \tag{2.15}$$

$$\pi_z = \frac{\sum_{i=1}^{N} P(z|\mathbf{\Omega_0}, \mathbf{\Theta}, \boldsymbol{\pi})}{N} \tag{2.16}$$

The two steps are repeated until convergence is achieved. The EM algorithm restores the additive separability at the maximization step in equation (2.15). Therefore estimation can be done sequentially. Arcidiacono and Jones (2003) name this procedure Expectation Sequential Maximization because it applies the EM algorithm to a sequential maximization problem. They show that this method produces consistent estimates of the parameters with large computational savings.

2.8 Results

2.8.1 The Dynamic Model

We first show the fit of the model and then present the estimated coefficients. A likelihood ratio test suggests that 3 types are sufficient to capture the unobserved heterogeneity: 60% of the individuals in our cohort are estimated to be type 1, 35% of type 2 and 5% of type 3. So far we have remained silent on what is unobserved heterogeneity. In structural schooling models like ours, researchers call it skill endowment (Keane and Wolpin (1997), Keane and Wolpin (2001)), or ability (Belzil and Hansen (2002) and Arcidiacono (2005)). There is also growing attention to the importance of non-cognitive skills as determinants of labor market outcomes and schooling (see Heckman et al. (2006)). Since in our model we include early measures of cognitive skills among the observed characteristics, we do not consider the unobserved heterogeneity as ability. Rather we prefer to think of unobserved heterogeneity as those non-cognitive skills that are uncorrelated with cognitive ones. Figure 2.4 shows the true and simulated educational choices. The model fits quite well the educational choices. From the figure we can see that type 2 are the more likely to go into education, followed by type 1. Table 2.4 reports the fit by education, work and unemployment status. The model fits less well the way individuals allocate themselves between work and unemployment. The reason for this latter result is the peak in the unemployment series in period 8, at the age of 22, which the model can not really fit given that the job offer (δ) and job firing (ϕ) probabilities are not period specific.

However we only model the staying-on choice. Having left the education system, work or unemployment are not choices. Moreover, our main focus is to predict the decision of leaving school, which our model does reasonably well excluding the case of A Levels enrolment where there is a 1.7% gap between true and predicted data.

Figure 2.5 shows the wage fit. The model predicts quite well the wages by education groups with the exception of the A Level case, where there is a 2800 pounds difference. Unfortunately, given the low enrolment rates in A Level and Higher Education, we do not have many wage observations for these qualification so the imperfect fit is not fully surprising. Table 2.5 shows the true and predicted wage by time and qualification. The average wage is quite well predicted for all the 3 age points available in the data. However, when we decompose it by qualification, our model slightly under-estimates the unconditional wage return to A Levels and HE at the age of 33 and 41.

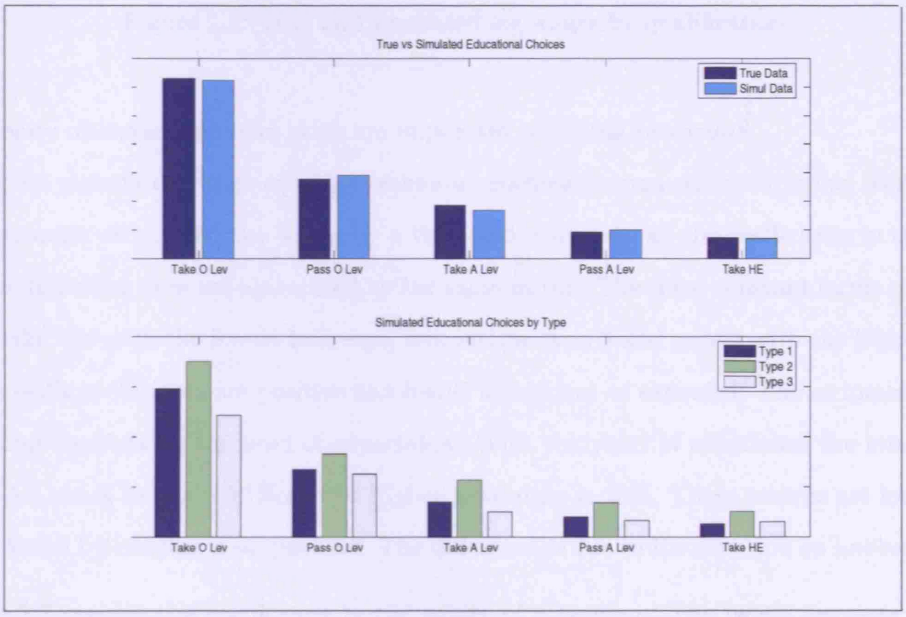

Figure 2.4: True and simulated educational choices

Figure 2.6 and 2.7 show how the model fits the educational outcomes by cognitive skill. In both cases the model matches fairly well the data, even though the simulation slightly over-estimates the educational outcomes of the lower quantiles. Even conditioning on unobserved

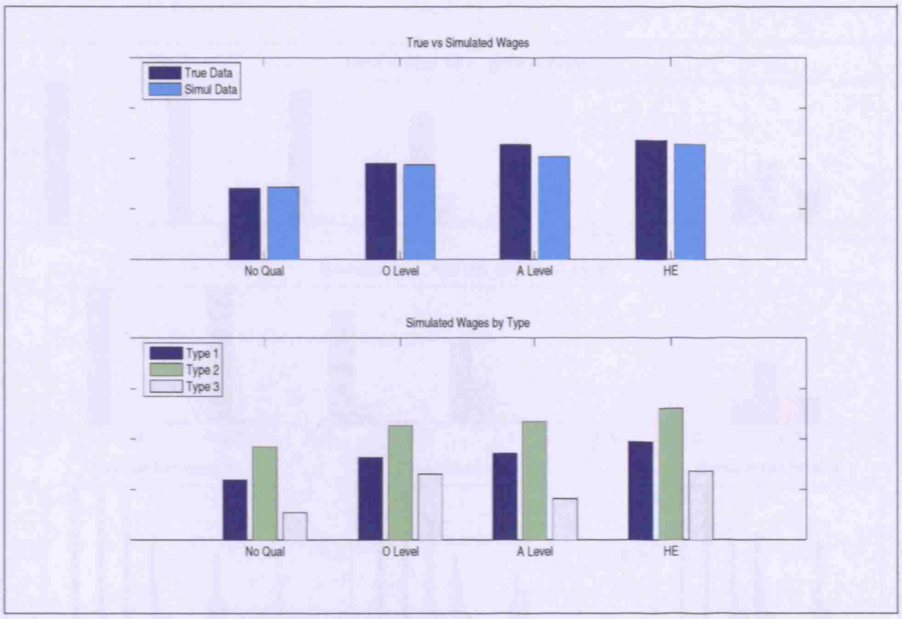

Figure 2.5: True and simulated log wages by qualification

heterogeneity observed cognitive skills are important schooling predictors.

Table 2.6 reports the wage equation estimated coefficients and standard errors. For computational reasons we divided the wages by a thousand, therefore all the coefficients in this table and in the following ones are also scaled by the same factor. The three constant terms show that type 1 ranks last with the lowest intercept, followed by type 3 and type 2 with the highest. The cognitive skills coefficients are positive and highly significant as expected. The estimated return to schooling depends on the level of experience. With ten years of experience the return to O Levels is 5%, to A Levels is 13% and to Higher Education is 30%. These returns are lower than what we found by running a simple OLS. The difference is due to the selection on unobservables.

Table 2.7 reports the coefficients in the utility of schooling. The model suggests that individuals had a negative utility from O Levels, with the utility being the highest for type 2. The negative utility is due to the relatively high return to O Levels and the absence of foregone earnings for this choice. Given free education, the only reason not to take O Levels must be their relatively high effort cost. The utility from A Levels and Higher Education are instead positive because of foregone earnings and its not surprising because education was totally free.

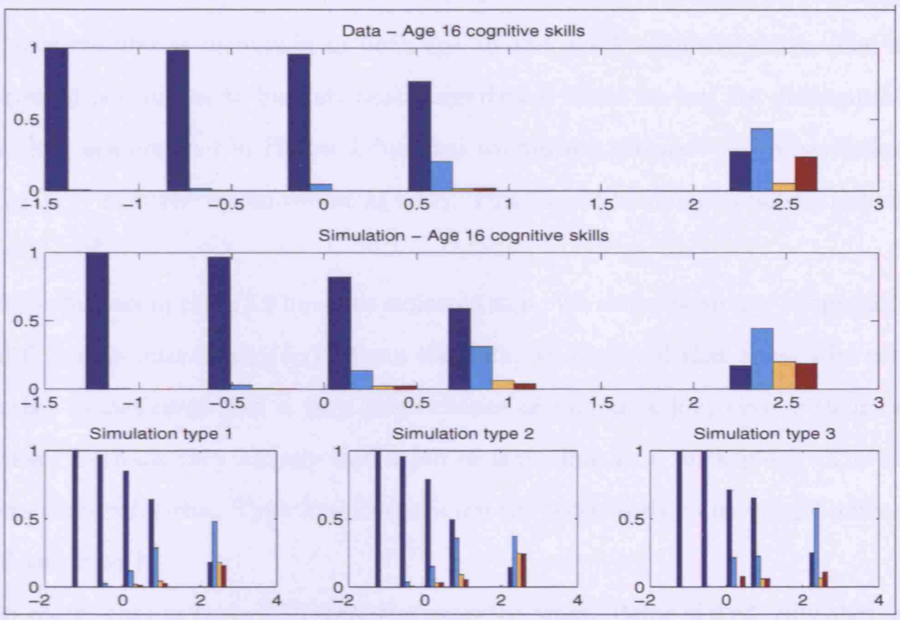

Figure 2.6: Qualifications by age 16 skills, true and simulated

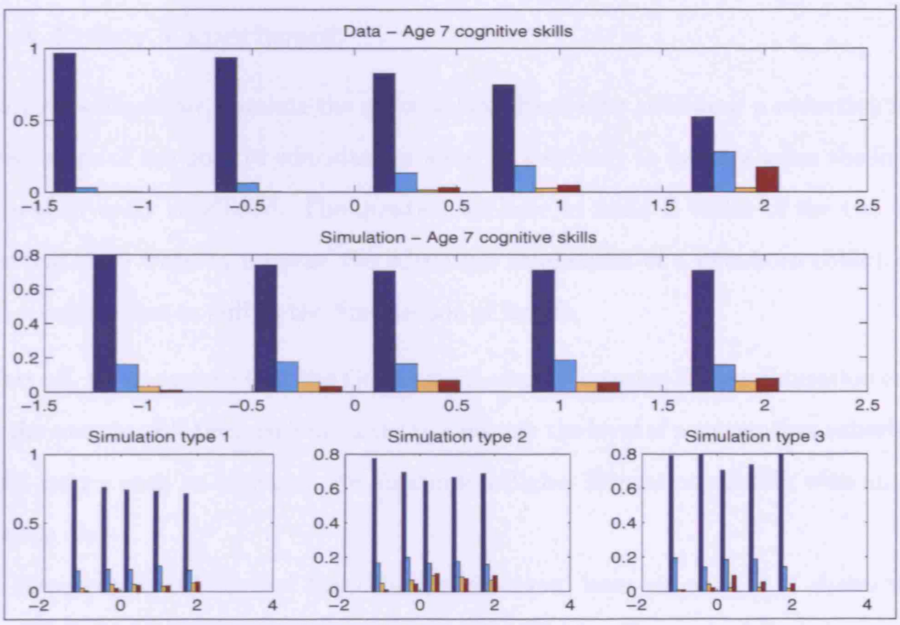

Figure 2.7: Qualifications by age 7 skills, true and simulated

The $\boldsymbol{\lambda}$ coefficients in table 2.8 indicate that the probability of success in obtaining a qualification once enrolled is increasing in both age 16 and age 7 cognitive skills. The difference across types do not appear to be statistically significant. Since we can not distinguish between enrolment and achievement in Higher Education we can not estimate λ_3. Nevertheless, rather than fixing $\lambda_3 = 1$, in the model we set $\lambda_3 = \lambda_2$. This should be a more realistic approximation of the true λ_3. [19]

The $\boldsymbol{\delta}$ coefficients in table 2.9 have the expected sign. We estimate an age 16 specific constant (δ_{01}) and O Levels interaction (δ_{11}). From the data, we observed that those who entered the labor market immediately had a very large chance of finding a job, even with no academic qualification. Perhaps they already had a job or some link to a work place. The additional parameters correct for this. Type 3 individuals are the most likely to have a job offer, followed by type 2 and type 1.

The $\boldsymbol{\phi}$ coefficients in table 2.10 show the expected sings. Being skilled, educated and experienced all reduce the probability of being fired. Again type 3 individuals are the least likely to be fired, followed by type 2 and type 1.

2.8.2 A Policy Experiment

As a policy experiment we simulate the effect of two alternative subsidies: a reduction in tuition fees, at any stage of the post 16 education process, or a subsidy to families when the individuals are still in their early childhood. The question we have in mind is which of the two would be more effective if we want to increase the education attainment of a new-born cohort or, given our data, a cohort that is still in the first decade of its life.

To start off, let us assume that the Government aims to increase Higher Education enrolment by 1% of the sample. We then use our model to compute the level of a tuition fees subsidy (grant) that would ensure such an increase. We alternate a Higher Education subsidy with an A Levels and O Levels ones.

With the exception of the last two columns, discussed later on, table 2.11 shows the result of performing this experiment. The upper part of the table reports the fraction of individuals

[19] Essentially we force individuals to believe there is a possibility of failing in Higher Education even though in the likelihood no one does.

enrolled in education at each age with or without subsidy. [20] We just report the results up to age 18 because this is when individuals enroll in higher education. The second part of the table reports respectively the amount of the subsidy, its ratio with the median parental income, and the per-capita cost. The latter takes into account the length and eligibility of each subsidy. An HE subsidy would be given for 3 years but only to those who enroll in HE. An AL subsidy would be given for 2 years to those who enroll in AL. An OL subsidy would be given once to those who enroll in OL.

Both an HE and AL subsidy succeeded in raising Higher Education enrolment by 1%, but the HE subsidy is more efficient. It is lower both in absolute terms and per-capita. The 1402 pounds subsidy correspond to 9.5% of the median parental income, costing 389 pounds each. We could not find an OL subsidy that would satisfy our initial requirement. The algorithm failed to find any improvement when it reached a subsidy worth more than 16824 pounds. For that value 86% of the individuals are enrolling in O Levels and yet the vast majority leave before Higher Education.

This result is clearly not surprising and embedded in the dynamic model. Subsidizing O Levels and A Levels does not change the incentives to get into Higher Education. The reason why some would enroll in Higher Education with OL and AL subsidies is that once the subsidy has brought them into O Levels or A Levels, they might get a large and positive taste shock for schooling and therefore stay even longer. Thus the main reason to test the effect of the grant at different educational levels is to show its overall effect. Even though these different subsidies are all raising HE enrolment by 1% (but in the OL case), they have different effects on AL and OL enrolment. As the second and third rows of table 2.11 show, the AL grant leads to higher AL enrolment than a HE grant. A similar argument applies to the OL grant. In defining the most efficient subsidy we are therefore assuming that the Government is only aiming for an increase in Higher Education while the externalities generated by higher OL and AL enrolment are negligible.

It is worth noting that Keane and Wolpin (1997) find a 8.4% increase in college graduation

[20] Note that even without subsidy the figures in table 2.11 are different from those in table 2.4. This is because table 2.4 takes into account sample attrition: i.e. sample attrition is changing the distribution of skills and parental income over age in our sample, and the simulation in table 2.4 corrects for that. In table 2.11 instead we keep the initial distribution of skills and parental income at age 16^a to simulate the whole process.

rates following a \$ 2000 grant in a similar set up, a much bigger effect that we find. It is also true that pre-grant college graduation rate in their sample is 24.2%, also much larger than our 8.2%. It is hard to say whether this difference is due to differences in the educational systems and education incentives between the US and the UK, to different years of birth or to the different way we model the choices.

Next, we try to test whether a parental income subsidy would be more efficient than the HE one in pursuing our 1% increase. Our hypothesis is that a parental income subsidy would increase the age 16 cognitive skill endowment, and in turn enrolment in education. In figure 2.6 we have seen that our model predicts that highly skilled individuals are more likely to enroll. There is no other direct effect of a parental income subsidy on educational choices, not even through the unobserved heterogeneity. Although this might appear restrictive, it is in line with the current literature findings that educational choices are affected by parental income only indirectly through the stock of skills.

We assume that our age 16 skill level is a function of innate ability μ and a history of parental background inputs PB:

$$K_{16} = f(\mu, PB_1, .., PB_{16}) \tag{2.17}$$

in particular, we are interested in disentangling the effect of parental income among the parental background factors. If we assume f to be linear in its arguments then:

$$E(K_{16}|Y_1, .., Y_{16}) = \boldsymbol{\eta} \boldsymbol{Y} \tag{2.18}$$

where $\boldsymbol{\eta}$ is a 1×16 vector of coefficients, and $\boldsymbol{Y}$ is a 16×1 vector of parental income inputs. Ideally we would like to know $max\{\boldsymbol{\eta}\}$ to subsidize parents at the most efficient point in time. We could think of a one off monetary subsidy, or voucher, that has to be spent within the year.

Unfortunately, our data do not provide us which such a long history of parental income: income is reported only at age 16. However, we have rich information on parental background characteristics at age 16, 11 and 7. These time varying variables would include father's social class, whether the mother was working, region of residence and of course age. Call $\boldsymbol{PBC}_t$ the vector including these variables at any time t.

We then try to infer parental income at a few points in time by first estimating a reduced form

$$Y_{16} = \boldsymbol{\zeta}_{16}\boldsymbol{PBC}_{16} + v \tag{2.19}$$

and, under the assumption that $\boldsymbol{\zeta}_{16} \approx \boldsymbol{\zeta}_{11} \approx \boldsymbol{\zeta}_7$, by computing

$$\hat{Y}_t = \hat{\boldsymbol{\zeta}}_{16}\boldsymbol{PBC}_t \tag{2.20}$$

for $t = 7, 11, 16$. Here $\hat{\boldsymbol{\zeta}}_{16}$ includes the constant and therefore the permanent income component.

Finally we take

$$E(K_{16}|\hat{Y}_7, \hat{Y}_{11}, \hat{Y}_{16}, K_7) = \boldsymbol{\eta}\hat{\boldsymbol{Y}} + \varsigma K_7 \tag{2.21}$$

where we include K_7 as a proxy of innate ability. A simple OLS estimation gives:

$$\eta_7 = 0.01; \quad \eta_{11} = 0.026; \quad \eta_{16} = 0.066; \tag{2.22}$$

all being statistically significant. [21]

There is no need to say that these estimates are quite rough approximations of the true $\boldsymbol{\eta}$'s. Upward bias is likely to be induced by omitted variables and downward bias by measurement error. The economic literature on the topic is still far from being able to clearly establish the true link between parental income and skills. [22] Dahl and Lochner (2005) use the Earning Tax Credit scheme as an instrument for parental income. They estimate that a 10000 dollars increase in income raises math test scores by 21% and reading test scores by 36% of a standard deviation. Their estimated is above what usually found in the literature using other instruments, fixed effects or simple OLS. [23] Our results suggest that an increase of 10000 pounds in 2001 prices

[21] Income is in 1000 pounds.

[22] See Todd and Wolpin (2003), Cunha and Heckman (2007) and Cunha, Heckman, Lochner, and Masterov (2006) for a discussion on the skills production function.

[23] Their data are the children of the NLSY79 cohort, aged 9.5 on average.

(around 15000 dollars), would push up the principal component by 50% of a standard deviation, close to their result. In any case, we do not feel in the position of claiming a consistent estimate of η either. We only have limited information on parental income, and we do not believe any of the variables at our disposal could be a valid instrument.

Therefore we use the model to compute the value of η for which a one-off parental income subsidy would be as powerful as the HE tuition fees subsidy. Given that the two subsidies would have to cost the same, we set the parental income subsidy to 389 pounds, the per-capita cost of the HE subsidy. The result is reported in the fourth column of table 2.11 (P. Inc. Sub (1)). We do manage to get an increase in HE enrolment by 1%, but this occurs for a value of $\tilde{\eta}$ equal to 0.216. This would be 3.24 times our estimated η_{16} implying that an increase of 10000 pounds would push up the score by 168% of a standard deviation. This is far above any estimate found in the skill literature, making it unlikely that even with a consistent estimate of η the parental income subsidy could have been as powerful as the HE one.

It is instructive to compare the age 16 cognitive skills distribution before and after the parental income subsidy. The mean skill after the parental subsidy is 6.5% of a standard deviation larger then before the subsidy. This implies that any policy aiming to achieve our +1% in enrolment only through cognitive skill accumulation would have to increase skills by at least 6.5% of a standard deviation.

Given the non-linearity of the schooling choice model, this result depends inevitably on the targeted raise in Higher Education enrolment. The ratio between $\tilde{\eta}$ and our estimated η_{16} monotonically decreases as we target a larger augment in Higher Education enrolment, going down to 2.15 when we move from a +1% to a +5% enrolment. However such an increase would likely have a strong effect on the return to education and cognitive skills, which our model does not account for. Heckman, Lochner, and Taber (1998) and Lee (2005) predicts a smaller impact of tuition subsidies once the general equilibrium effect, i.e. a lower return to schooling, is considered. Intuitively, if a general equilibrium effect was to appear in our model, the result of our comparison should be reinforced. While both kinds of subsidy would lead to a lower return to schooling, shifting the distribution of cognitive skills towards the right would also lower its return and therefore weaken the positive link between cognitive skill and education.

Thus, a larger increase in cognitive skills would be needed to reach the targeted increase in Higher Education enrolment.

Next, we check what would happen if rather than targeting the age 16 cognitive skills stock, we were to target the age 7 one. If cognitive skills are shaped early in life, and if age 7 cognitive skills have an effect on schooling choices even conditioning on age 16 ones, as the coefficients indicate, than targeting the age 7 skills stock might be more efficient. This assumes that K_7 is not exogenous and therefore not a good proxy of innate ability. The ς coefficient in the production function, equation (2.21), would be inconsistent and likely upward biased. Nevertheless, keeping this in mind, it is interesting to ask what change in age 7 cognitive skills is needed to match the effect of a 389 pounds Higher Education subsidy. We actually fall below the target by 0.86 percentage points since the effect of age 7 skills on educational choices is not strong enough. However, when the algorithm stops the return to parental income is already 0.21, that is 3.15 our OLS estimate. If $\hat{\varsigma}$ is upward biased then $\tilde{\eta}$ is a lower bound.

We also try to set a +1% target for A Level and O Level achievement. The Government might just value raising education to post-compulsory levels without necessarily increasing Higher Education enrolment. The results are shown in tables 2.12 and 2.13. As we would expect, the best way to reach the target is to give a grant at the target level, that is subsidize A Levels if the target is +1% enrolment in A Level, or similarly for O Levels. The parental income subsidy still require very high returns to be efficient, respectively 3.03 (A Levels) and 2.40 (O levels) times our OLS estimate when targeting age 16 skills.

A few more points worth noting. One could argue that this policy experiment is not very informative given that this cohort was born in 1958 and participation in Higher Education is much higher for the current cohorts. The importance of cognitive skills might have changed over time. Yet higher participation, either due to larger supply or demand, is likely to decrease the importance of cognitive skills, given that a now less elitarian group of individuals staying on. Therefore, our model is likely to over-estimate the impact of cognitive skills and, as a consequence, of the parental income subsidy, strengthening our conclusions.

The linear utility implies no borrowing constraints and no risk aversion. If the individuals did face a borrowing constraint then tuition fees subsidies (grants) would be even more effective than our model predicts. Concerning risk aversion, the only source of randomness in our model is the taste shock in the utility of schooling. There is no a priori reason why a tuition subsidy or an increase in cognitive skills should have very different effects due to risk aversion.

2.9 Conclusions

In this paper we build a dynamic structural model of educational choices and exploit rich cohort data to investigate the importance of observed cognitive skills, financial incentives and unobserved heterogeneity in educational choices. The model predicts that highly skilled individuals stay longer in education and it does well in replicating the choices conditional on the skill group.

We then simulate the effect of two educational subsidies equal in cost but different in timing. The first consists of grants assigned directly to the individuals aged between 16 and 18. The second is a subsidy assigned to the parents earlier on, when the cohort is still in its childhood. Our aim is to test whether an unconditional parental subsidy, with its indirect effect on cognitive skills, could be more efficient in fostering educational enrolment than a direct educational grant. When we target an increase in Higher Education enrolment equal to 1% of the population, our results suggest that this is not the case. Unless the effect of parental income on skills is implausibly high, a direct grant in the form of a tuition subsidy seems more efficient even in the absence of short term financial constraints. Although cognitive skills accumulated during childhood play a key role in the educational decisions, an unconditional financial subsidy to parents is not the best policy. This result is robust even if we target a different increase in Higher Education enrolment or if we target an increase at a lower education level.

The overall result does not call a halt to additional investment in cognitive skill accumulation during childhood. We only consider the link between cognitive skills and parental income, surely weaker than between skills and parental investment, such as child care or parental time. In our opinion, the insight of the paper is the opposite, cognitive skills are important but changes in educational choices demand a sizeable shift in the cognitive skills distribution.

Table 2.1: Status by age (Number of observations)

Age	Education	Work	Unemployment	Total
16[a]	1,811	0	1,063	2,874
	(63.01)	(0.00)	(36.99)	(100.00)
16[b]	535	2,178	161	2,874
	(18.62)	(75.78)	(5.60)	(100.00)
17	535	2,222	109	2,866
	(18.67)	(77.53)	(3.80)	(100.00)
18	197	2,349	153	2,699
	(7.30)	(87.03)	(5.67)	(100.00)
19	197	2,341	129	2,667
	(7.39)	(87.78)	(4.84)	(100.00)
20	197	2,318	129	2,644
	(7.45)	(87.67)	(4.88)	(100.00)
21	0	2,407	153	2,560
	(0.00)	(94.02)	(5.98)	(100.00)
22	0	2,208	267	2,475
	(0.00)	(89.21)	(10.79)	(100.00)
23	0	2,190	252	2,442
	(0.00)	(89.68)	(10.32)	(100.00)
24	0	1,729	120	1,849
	(0.00)	(93.51)	(6.49)	(100.00)
25	0	1,714	119	1,833
	(0.00)	(93.51)	(6.49)	(100.00)
26	0	1,692	123	1,815
	(0.00)	(93.22)	(6.78)	(100.00)
27	0	1,687	113	1,800
	(0.00)	(93.72)	(6.28)	(100.00)
28	0	1,692	100	1,792
	(0.00)	(94.42)	(5.58)	(100.00)
29	0	1,686	92	1,778
	(0.00)	(94.83)	(5.17)	(100.00)
30	0	1,686	81	1,767
	(0.00)	(95.42)	(4.58)	(100.00)
31	0	1,676	83	1,759
	(0.00)	(95.28)	(4.72)	(100.00)
32	0	1,654	101	1,755
	(0.00)	(94.25)	(5.75)	(100.00)
33	0	1,458	81	1,539
	(0.00)	(94.74)	(5.26)	(100.00)
34	0	1,413	78	1,491
	(0.00)	(94.77)	(5.23)	(100.00)
35	0	1,407	76	1,483
	(0.00)	(94.88)	(5.12)	(100.00)
36	0	1,400	78	1,478
	(0.00)	(94.72)	(5.28)	(100.00)
37	0	1,391	82	1,473
	(0.00)	(94.43)	(5.57)	(100.00)
38	0	1,381	85	1,466
	(0.00)	(94.20)	(5.80)	(100.00)
39	0	1,372	87	1,459
	(0.00)	(94.04)	(5.96)	(100.00)
40	0	1,360	90	1,450
	(0.00)	(93.79)	(6.21)	(100.00)
41	0	1,332	108	1,440
	(0.00)	(92.50)	(7.50)	(100.00)
Total	4,047	45,943	4,251	54,241
	(7.46)	(84.70)	(7.84)	(100.00)

Table 2.2: Wage (Sample averages)

	age=23	age=33	age=41
Wage	12326.47	19175.00	23904.71
	(3220.47)	(7209.19)	(11409.47)
	Wage by qualification.		
Wage — No Qualif.	12094.48	17607.02	21381.15
	(3233.81)	(6115.98)	(9011.91)
Wage — O Levels	13125.45	22878.02	28981.42
	(3050.66)	(7750.02)	(12156.04)
Wage — A Levels	13415.03	25742.56	37967.17
	(2670.46)	(6602.89)	(18981.43)
Wage — H.E.	13403.48	28595.93	38780.83
	(3022.19)	(7964.21)	(15630.93)

Standard Deviations in brackets.
Yearly wages are derived by first computing hourly wages and then multiplying them by 52. Wages are in January 2001 prices.

Table 2.3: Cognitive Skills and Parental Income (Sample averages)

	age 16	age 7
Math Score	13.68	5.33
	(7.26)	(2.44)
Reading Score	25.70	23.15
	(7.10)	(7.00)
Principal Component	-2.28e-10	-2.34e-09
	(1.28)	(1.17)
Parental Income	16226.79	—
	(6128.49)	(—)

Standard Deviations in brackets.

Table 2.4: Model Fit - Status (Proportions)

Age	Education True	Education Predicted	Work True	Work Predicted	Unemp. True	Unemp. Predicted
16[a]	0.6301	0.6241	0	0	0.3699	0.3759
16[b]	0.1862	0.1688	0.7578	0.7805	0.0560	0.0507
17	0.1867	0.1675	0.7753	0.7850	0.0380	0.0475
18	0.0730	0.0701	0.8703	0.8688	0.0567	0.0611
19	0.0739	0.0672	0.8778	0.8794	0.0484	0.0534
20	0.0745	0.0657	0.8767	0.8798	0.0488	0.0545
21	0	0	0.9402	0.9442	0.0598	0.0558
22	0	0	0.8921	0.9436	0.1079	0.0564
23	0	0	0.8968	0.9418	0.1032	0.0582
24	0	0	0.9351	0.9422	0.0649	0.0578
25	0	0	0.9351	0.9431	0.0649	0.0569
26	0	0	0.9322	0.9426	0.0678	0.0574
27	0	0	0.9372	0.9430	0.0628	0.0570
28	0	0	0.9442	0.9437	0.0558	0.0563
29	0	0	0.9483	0.9454	0.0517	0.0546
30	0	0	0.9542	0.9451	0.0458	0.0549
31	0	0	0.9528	0.9452	0.0472	0.0548
32	0	0	0.9425	0.9448	0.0575	0.0552
33	0	0	0.9474	0.9465	0.0526	0.0535
34	0	0	0.9477	0.9473	0.0523	0.0527
35	0	0	0.9488	0.9475	0.0512	0.0525
36	0	0	0.9472	0.9481	0.0528	0.0519
37	0	0	0.9443	0.9490	0.0557	0.0510
38	0	0	0.9420	0.9484	0.0580	0.0516
39	0	0	0.9404	0.9471	0.0596	0.0529
40	0	0	0.9379	0.9460	0.0621	0.0540
41	0	0	0.9250	0.9443	0.0750	0.0557

Table 2.5: Model Fit - Wage (Sample average)

	True	Predicted	True	Predicted	True	Predicted	True	Predicted
	Overall		Age 23		Age 33		Age 41	
	17750	18003	12326	12635	19175	18094	23905	23518
No Qualif	16381	16590	12094	12338	17607	16749	21381	20866
O Levels	20987	20722	13125	13463	22878	20756	28981	28299
A Levels	25297	22413	13415	13061	25743	22089	37967	32532
HE	26341	25299	13403	13797	28596	24786	38781	37809

Table 2.6: Estimated Coefficients - Wage Equation

	Estimate	S.E.	ratio
σ_u	0.2789	(0.0027)	(104.5020)
Type 1	2.2286	(0.0147)	(151.1216)
Type 2 deviation from type 1	0.3229	(0.0118)	(27.3637)
Type 3 deviation from type 1	0.0249	(0.0644)	(0.3862)
α_{k2}	0.0654	(0.0073)	(8.9546)
α_{k1}	0.0190	(0.0056)	(3.4003)
α_{q1}	-0.0614	(0.0356)	(-1.7277)
α_{q2}	-0.0676	(0.1227)	(-0.5511)
α_{q3}	0.0354	(0.0559)	(0.6331)
α_x	0.0278	(6.7610e-004)	(41.0976)
α_{x1}	0.0118	(0.0017)	(6.7742)
α_{x2}	0.0202	(0.0060)	(3.3557)
α_{x3}	0.0271	(0.0030)	(9.0672)

Table 2.7: Estimated Coefficients - Utility of Schooling

	Estimate	S.E.	ratio
σ_{ϵ_1}	18.0527	(2.5486)	(7.0834)
σ_{ϵ_2}	14.9309	(4.1329)	(3.6127)
σ_{ϵ_3}	21.9125	(9.3737)	(2.3376)
Type 1	-5.5077	(1.1841)	(-4.6512)
Type 2 deviation from type 1	4.1302	(1.6581)	(2.4909)
Type 3 deviation from type 1	2.6691	(2.4009)	(1.1117)
γ_{AL}	12.0880	(3.2117)	(3.7637)
γ_{HE}	11.8215	(3.3493)	(3.5295)

Table 2.8: Estimated Coefficients - Probability of Success in School

	Estimate	S.E.	ratio
Type 1	-0.9936	(0.0816)	(-12.1709)
Type 2 deviation from type 1	0.0754	(0.1477)	(0.5108)
Type 3 deviation from type 1	-0.0412	(0.2444)	(-0.1687)
λ_{1k2}	1.3638	(0.0371)	(36.7728)
λ_{1k1}	0.1081	(0.0382)	(2.8314)
λ_{AL}	0.5481	(0.1631)	(3.3613)
λ_{2k2}	0.4946	(0.0857)	(5.7715)
λ_{2k1}	0.1357	(0.0387)	(3.5074)

Table 2.9: Estimated Coefficients - Probability of having a Job Offer

	Estimate	S.E.	ratio
Type 1	0.4841	(0.0674)	(7.1855)
Type 2 deviation from type 1	0.2103	(0.1264)	(1.6633)
Type 3 deviation from type 1	1.3726	(0.0599)	(22.9256)
δ_{k2}	0.1211	(0.0305)	(3.9739)
δ_{k1}	0.0054	(0.0258)	(0.2090)
δ_{q1}	0.0986	(0.0911)	(1.0829)
δ_{q2}	0.0229	(0.2003)	(0.1145)
δ_{q3}	0.9012	(0.1520)	(5.9292)
δ_{x}	-0.0741	(0.0050)	(-14.8447)
δ_{01}	1.2630	(0.0726)	(17.4067)
δ_{11}	-0.7956	(0.1354)	(-5.8780)

Table 2.10: Estimated Coefficients - Probability of being fired

	Estimate	S.E.	ratio
Type 1	-1.7899	(0.0335)	(-53.4109)
Type 2 deviation from type 1	-0.6224	(0.0836)	(-7.4437)
Type 3 deviation from type 1	-1.0765	(0.0778)	(-13.8312)
ϕ_{k2}	-0.1378	(0.0198)	(-6.9570)
ϕ_{k1}	0.0034	(0.0159)	(0.2120)
ϕ_{q1}	-0.1980	(0.0588)	(-3.3684)
ϕ_{q2}	0.0741	(0.1616)	(0.4585)
ϕ_{q3}	0.0526	(0.1138)	(0.4619)
ϕ_{x}	-0.0200	(0.0024)	(-8.4790)

Table 2.11: Policy Experiment - +1% in HE enrolment

	No Sub	O Levels Sub.	A Levels Sub.	HE Sub	P. Inc. Sub (1)	P. Inc. Sub (2)
16[a]	0.6241	0.8611	0.6278	0.6255	0.6363	0.6326
16[b]	0.1642	0.1747	0.1876	0.1717	0.1773	0.1712
17	0.1642	0.1747	0.1876	0.1717	0.1773	0.1712
18	0.0825	0.0852	0.0925	0.0925	0.0925	0.0839
Subsidy	0	16824	1932	1402	1402	1402
% of Median Income	0	1.1306	0.129	0.094	0.094	0.094
Per-Capita Cost	0	14487	725	389	389	389

Table 2.12: Policy Experiment - +1% in AL enrolment

	No Sub	O Levels Sub.	A Levels Sub.	HE Sub	P. Inc. Sub (1)	P. Inc. Sub (2)
16[a]	0.6241	0.8779	0.6259	0.6256	0.6328	0.6305
16[b]	0.1642	0.1742	0.1742	0.1742	0.1742	0.1724
17	0.1642	0.1742	0.1742	0.1742	0.1742	0.1724
18	0.0825	0.0849	0.0870	0.0949	0.0912	0.0866
Subsidy	0	18529	877	1724	877	877
% of Median Income	0	1.245	0.059	0.115	0.059	0.059
Per-Capita Cost	0	16267	305	490	305	305

Table 2.13: Policy Experiment - +1% in OL enrolment

	No Sub	O Levels Sub.	A Levels Sub.	HE Sub	P. Inc. Sub (1)	P. Inc. Sub (2)
16[a]	0.6241	0.6341	0.6341	0.6341	0.6341	0.6335
16[b]	0.1642	0.1653	0.2158	0.2187	0.1747	0.1672
17	0.1642	0.1653	0.2158	0.2187	0.1747	0.1672
18	0.0825	0.0827	0.1054	0.1423	0.0863	0.0857
Subsidy	0	689	4319	8563	689	689
% of Median Income	0	0.046	0.290	0.575	0.046	0.046
Per-Capita Cost	0	437	1864	3655	437	437

Chapter 3

Testing Stochastic Monotonicity in Intergenerational Mobility Tables

3.1 Introduction

The extent to which individuals inherit their socio-economic status in a society has important implications for policy debates on equal opportunities and social justice. The analysis of the causes and consequences of the intergenerational transmission of social status from one generation to the next has traditionally been one of the most active areas of research by economists, sociologists and social scientists alike.

Inspection of typical mobility tables and theoretical reasoning indicate that in most societies there is a general tendency for children from higher status parents to somehow fare "better" in social achievement than children from lower status parents. To substantiate this general idea, first, one needs to agree on the exact meaning of this statement, second, one has to devise a sort of testing procedure for verifying the hypothesis, and finally, one has to apply the testing procedure with real world data.

In this paper we will consider this characteristic of the intergenerational transmission of social status, which implicitly or explicitly underlines most research on intergenerational social mobility. The first section of this paper will propose a precise definition of the idea that a child is better off by having a parent with a higher social status, based on the theory of monotone Markov chains as developed by Keilson and Kester (1977) and Conlisk (1990). We then employ a rich data set on intergenerational social mobility, which has been made available in Ganzeboom, Luijkx, and Treiman (1989), to test the monotonicity assumption using the stochastic dominance testing procedure of Dardanoni and Forcina (1998). This large data set contains 149 mobility tables combining information from 35 different countries and different years. It is a very comprehensive data set for comparative mobility analysis, and it has the distinctive advantage of employing a consistent and well defined definition of social status. Perhaps not surprisingly, we find that for most societies the monotonicity assumption cannot be rejected.

In the second part of the paper, we consider testing for stochastic monotonicity conditional on some appropriate set of covariates z. Starting with Becker and Tomes (1979), researchers have proposed economic models of intergenerational mobility. It is by now widely accepted that parental transmission of skills, beliefs, motivation and social connections are all important in explaining the strong dependence between father and son social status. It is therefore natural to test whether the positive dependence between father's and son's status is still present after conditioning for some of these characteristics. On a similar line of thought, some researchers have turned their attention to the concept of equality of opportunity (*EoP*).[1] Dardanoni, Fields, Roemer, and Puerta (2006) for example, following the seminal contribution of Putterman, Roemer, and Silvestre (1998) and Roemer (2000), describe *EoP* by distinguishing between circumstances and effort. Circumstances are aspects of the environment affecting the socio-economic status and for which society does not wish to hold individuals responsible. Effort is the set of actions affecting the status for which individuals are responsible. *EoP* implies that differences in status are ethically acceptable if they are due to differential effort but not if they are due to differential circumstances. This requires independence conditional on those covariates z that we consider effort.

Extending the linear regression model to control for some covariates z is straightforward. However, the transition matrix literature has reported indices of mobility based on unconditional distributions, or at most conditional on a few discrete covariates. The main theoretical challenge is to devise statistical inference procedures to test for stochastic monotonicity conditional on observed covariates. To this purpose, we extend the Dardanoni and Forcina (1998) test for stochastic monotonicity by allowing for conditioning on covariates. Technically, our approach employs recent advances in marginal modelling (see e.g. Bergsma and Rudas (2002)

[1] Defining the appropriate concepts of equality of opportunity and testing them empirically is an area which is undergoing much current research, see e.g. Bourguignon, Ferreira, and Menendez (2003), Peragine (2004), Lefranc, Pistolesi, and Trannoy (2006) and Fleurbaey (2008).

and Bartolucci, Colombi, and Forcina (2007)), along the lines of Bartolucci, Forcina, and Dardanoni (2001), who consider (unconditional) testing for a notion of positive dependence (*positive quadrant dependence*) which is weaker than the monotonicity assumption analyzed in this paper.

We apply our methodology using the National Child Development Survey (NCDS), a UK cohort data set, which follows a cohort born in 1958 over its lifetime. Information on social class and wages is available both for the cohort members and their parents. The data also provide information on the educational attainment, cognitive and non-cognitive skills of the cohort members. Given the amount of characteristics observable to the researcher, the data are particulary fit to test for conditional dependence and different notions of equality of opportunity. Our results indicate that even though our control variables explain part of the intergenerational mobility mechanism, stochastic monotonicity holds both unconditionally and conditionally.

3.2 Monotone transition matrices

Let X and Y denote, respectively, father's and son's lifetime socio-economic status, and assume they take k possible values, which correspond to the k socio-economic classes that are ordered from best to worst. Consider then a standard discrete Markov chain of intergenerational social mobility: if the unit of observation is a family of father and child, the chain can be described by the equation $\boldsymbol{p}^{y\prime} = \boldsymbol{p}^{x\prime}\ \boldsymbol{P}$, where $\boldsymbol{P}$ denotes the $(k \times k)$ transition matrix, with typical element $P_{ij} = \Pr(son\ in\ j \mid father\ in\ i) \geq 0$, and $\boldsymbol{p}^x$, $\boldsymbol{p}^y$ denote respectively the marginal distributions of father's and son's social status. The typical row i of an intergenerational transition matrix indicates the probability distribution faced by a son whose father belongs to social class i. As argued above, it is natural to expect that when social states are ordered, sons whose fathers are in a higher social class are somewhat at an advantage with respect to the sons whose fathers are in a lower class. In a stochastic setting, this translates into the assumption that the "lottery" faced by the son of a father in class i is better than the "lottery" faced by the son of a father in class $i+1$. A natural definition of a "better lottery" in this context is given by the stochastic dominance ordering $\succeq$: given two $(k \times 1)$ probability vectors $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$, we say that $\boldsymbol{q}_1 \succeq \boldsymbol{q}_2$ if $q_{11} + q_{12} + \cdots + q_{1l} \geq q_{21} + q_{22} + \cdots + q_{2l}$ for all $l = 1, 2, \cdots, k-1$.

The stochastic dominance ordering gives a precise definition to the intuitive notion of differential advantage. Let $\boldsymbol{s}$ denote a real valued $(k \times 1)$ vector of "social status scores", where the typical element s_l reflects a quantitative measure of the value of social class l, and let $\boldsymbol{p}_i$ (a row vector) denote the ith row of $\boldsymbol{P}$. A son whose father is in class i will have with probability p_{il} social status s_l, so that his *expected social score* is given by $\boldsymbol{p}_i\boldsymbol{s}$. An equivalent characterization of the stochastic dominance ordering is obtained in terms of expected social status: $\boldsymbol{p}_i\boldsymbol{s} \geq \boldsymbol{p}_j\boldsymbol{s}$ (here and hereafter, when an inequality symbol involves vectors and matrices, we mean that the inequality is satisfied elementwise) for any increasing vectors $\boldsymbol{s}$ if and only if $\boldsymbol{p}_i \succeq \boldsymbol{p}_j$.[2] The intuitive notion of background advantage is captured in the discrete Markov chain by the so called 'monotonicity' assumption. The transition matrix of a discrete Markov chain with ordered states is called monotone (see Keilson and Kester (1977), Conlisk (1990), Dardanoni (1993) and Dardanoni (1995) for applications) if each row stochastically dominates the row below it, $\boldsymbol{p}_1 \succeq \boldsymbol{p}_2 \succeq \cdots \succeq \boldsymbol{p}_k$. It follows that in a monotone mobility process each son at time t is better off, in terms of expected social status, by having a parent from state i than by having a parent from state $i+1$.[3]

A key monotone transition matrix is the so-called "equal opportunities" transition matrix (see e.g. Prais (1955)), where at time t each son faces an identical probability distribution regardless of his father's background. Given the transition equation $\boldsymbol{p}^{y'} = \boldsymbol{p}^{x'}\boldsymbol{P}$, the equal opportunities mobility matrix is equal to $\mathbf{1}\boldsymbol{p}^{y'}$, so that the stochastic dominance constraint is satisfied as an equality. This particular monotone mobility matrix will play an important role in the hypothesis testing of the monotonicity assumption.

3.3 Testing unconditional monotonicity: An application

There is now an extensive statistical literature (see e.g. Robertson, Wright, and Dykstra (1988) and Silvapulle and Sen (2005)), on estimation and hypothesis testing in problems involving stochastic orderings. In particular, Robertson and Wright (1982) derive testing procedures based on maximum likelihood estimates of two stochastically ordered distributions, and Dykstra, Kochar, and Robertson (1991) obtain the maximum likelihood estimates of several multinomial

[2]This is a well known result in the stochastic dominance literature; see e.g. Lehmann (1955).

[3]Under the assumption of a constant transition matrix, this relationship also holds for the grandfather, grand-grandfather and so on. This follows from the fact that if $\boldsymbol{P}$ is monotone, so is $\boldsymbol{P}^t$ for $t = 1, 2, \cdots$.

distributions under uniform stochastic ordering, which is stronger than stochastic dominance. Dardanoni and Forcina (1998) extend these results and propose a nonparametric test for stochastic dominance which can be used to test monotonicity of the Markov chain of intergenerational mobility. In particular, Theorem 2 in their paper gives conservative bounds to the distribution of the likelihood ratio test statistic for testing monotonicity against an unrestricted alternative.

We perform Dardanoni and Forcina's procedure on a sample of cross-classification tables presented in Ganzeboom et al. (1989). This data set, which contains 149 intergenerational class mobility tables from 35 countries, is one of the most comprehensive and well structured data set on intergenerational social mobility to date. Ganzeboom, Luijks and Treiman present the cross-classification of father's occupation by son's current occupation for representative national samples of men aged 21-64, with the characteristic that the tables conform to a well specified six category scheme. The six social classes, in descending order of socio-economic status, are the following: 1) Large proprietors, higher and lower professionals and managers; 2) Routine non-manual workers; 3) Small proprietors with and without employees; 4) Lower grade technicians, manual supervisors and skilled manual workers; 5) Unskilled and semiskilled manual workers; 6) Self employed farmers and (unskilled) agricultural workers. The use of a common and well structured classification of social classes results in a substantial degree of comparability among the different tables.

Maximum likelihood estimation of each mobility matrix subject to the monotonicity constraint has been carried out by the iterative quadratic programming algorithm described by Dardanoni and Forcina (1998). Convergence to the fifth decimal place of the likelihood function is usually obtained within the first three iterations. The value of the likelihood ratio test statistic is reported in table 3.11 in the appendix, along with the number of monotonicity constraints actually binding in the sample.

Following Dardanoni and Forcina (1998, Theorem 2), the α critical values of the conservative unconditional chi-bar-squared test can be found by solving the equation

$$\sum_{i=0}^{(k-1)^2} \left(\begin{array}{c} (k-1)^2 \\ i \end{array} \right) 2^{-(k-1)^2} Pr[\chi_i^2 = c] = \alpha.$$

By numerical integration we obtain a value $c = 22.78$ at the 95% significance level, while it equals 28.61 at the 99% level. Comparing these significance levels with the value of the likelihood ratio test statistic, it emerges that out of 149 intergenerational class mobility tables, monotonicity is rejected at the 99% significance level only for the transition matrices of Hungary 1963, Philippines 1968, Poland 1972 and Spain 1975. In addition, the monotonicity hypothesis is rejected at the 95% level for Hungary 1973 and 1983 and India 1963c. Thus, it appears that monotonicity of the intergenerational transmission mechanism can generally be considered as an assumption supported by the real world.

3.4 Conditional Stochastic Monotonicity

A large number of studies have investigated the degree of intergenerational mobility. Irrespective of country, time period, measure of socio-economic status and measure of association, almost all studies have found that parent's and offspring's adult status are not independent, but exhibit some form of positive dependence. The previous section has confirmed this "fact of life", where positive dependence is precisely formulated in terms of stochastic dominance, and formal statistical inference procedures have been employed.

Starting with Becker and Tomes (1979) researchers have proposed economic models of intergenerational mobility to uncover the mechanism behind the transmission of social status. Becker and Tomes (1986), Solon (1999), Mulligan (1999), Han and Mulligan (2001) and Restuccia and Urrutia (2004) are all attempts in that direction. At a basic level, a simple model that assumes intergenerational transmission of ability and a human capital return to parental investment that is increasing in the child's ability can already generate a high degree of immobility. Adding imperfect capital markets to the model results in even less mobility. These models also show that the degree of mobility can be highly non-linear across the father and child's socio-economic distribution, whether socio-economic status is measured by wage, consumption or education. On a more intuitive ground, Bowles and Gintis (2002), Erikson and Goldthorpe (2002) and Blanden, Gregg, and Macmillan (2007) suggest that more than a simple transmission of ability might be in place. Other factors such as race, geographical location, wealth, risk aversion, discounting of the future, non-cognitive skills, but also height and beauty can be transmitted and generate the correlation in status. Not surprisingly most of these papers have also tried to empirically investigate the black box. However, none of these studies can explain more than 60% of the overall correlation. Bjorklund, Lindahl, and Plug (2006) use unique Swedish data with information on adopted children's biological and adoptive parents to estimate intergenerational mobility associations in earnings and education. They find that both pre- and post-birth factors contribute to intergenerational earnings and education transmission. The distinction between nature and nurture is particularly important if we are asked to judge meritocracy and equality of opportunity.

On a slightly different line of thought Dardanoni et al. (2006) discuss different notions of equality of opportunity based on the distinction between circumstances and effort: "Agreement is widespread that equality of opportunity holds in a society if the chances that individuals have to succeed depend only on their own efforts and not upon extraneous circumstances that may inhibit or expand those chances. What is contentious, however, is what constitutes effort and circumstances". In their paper they describe four channels through which parents affect status in an intergenerational context: social connections, the formation of social beliefs and skills, the transmission of native ability and the instillation of preferences and aspirations. Various notions of *EoP* depend on whether these channels are regarded as circumstances or effort. In other words, if we consider all those channels as circumstances out of an individual's control, than *EoP* implies perfect intergenerational mobility. This is perhaps the strongest definition of *EoP* where parent's and offspring's status must be independent. Less stringent notions of *EoP* allow for some of those channels to be influenced by the offsprings. In turn this requires independence conditional on those covariates z that we consider individual effort. z could include measures of preferences and aspirations, native ability, social beliefs and skills, and social connections.

3.5 Testing conditional monotonicity: Theory

Recall that X and Y denote, respectively, father's and son's social class, and let z be a vector of covariates which may affect the joint distribution and denote with $\pi(z)$ the vector containing the joint distribution of X and Y conditional on z, with Y categories running faster. If z was discrete and a sufficient number of observations were available for each distinct configuration of z, Dardanoni and Forcina (1998)'s unconditional test procedure could be performed for each subpopulation. However these conditions are unlikely to be satisfied whenever the number of covariates is reasonably large and/or certain covariates assume a large number of distinct values. Therefore, a meaningful approach is to model the effect of covariates by a suitable link function and a regression model.

The link function that we propose is based on the mapping of the conditional distribution of $X, Y \mid z$ into a set of row and column marginal parameters and $(k-1)^2$ association parameters. Both the row and column marginal parameters are *Global Logits* (see e.g. Agresti (2002)):

$$\rho_i(z) = \log\left[\frac{P(X > i \mid z)}{P(X \leq i \mid z)}\right], \; i = 1, \ldots, k-1;$$

$$\xi_j(z) = \log\left[\frac{P(Y > j \mid z)}{P(Y \leq j \mid z)}\right], \; j = 1, \ldots, k-1;$$

while the association parameters are *Local-Global Log-Odds Ratios* (see e.g. Agresti (2002)):

$$\tau_{ij}(z) = \log\left[\frac{P(X = i, Y \leq j \mid z)P(X = i+1, Y > j \mid z)}{P(X = i, Y > j \mid z)P(X = i+1, Y \leq j \mid z)}\right], \; i, j = 1, \ldots, k-1.$$

Douglas, Fienberg, Lee, Sampson, and Whitaker (1990) show that $\pi(z)$ is monotone if and only if the set of $(k-1)^2$ LG-log odds ratios are nonnegative. Global logits can be seen as the natural generalization of the standard binary logits when the variable is ordered; in fact, global logits can be interpreted as binary logits computed on successive splits of the ordinal response categories into "low" and "high" levels.

Collect now the corresponding parameters into the vectors $\rho(z), \xi(z)$ and $\tau(z)$ (by letting the j index run faster than i), then let $\lambda(z) = [\rho(z)', \xi(z)', \tau(z)']'$; this has dimension $2(k-1) + (k-1)^2 = k^2 - 1$, which equals the number of free parameters in $\pi(z)$. The results of Bartolucci et al. (2007) on marginal parameterizations may be applied to this context and imply that there exists a matrix of row contrasts C and a matrix M of zeros and ones such that, $\forall z$

$$\lambda(z) = C \log[M\pi(z)]$$

and the mapping from $\pi(z)$ to $\lambda(z)$ is invertible and differentiable for all strictly positive $\pi(z)$

(see their Theorem 1). Thus, the set of marginal and association parameters $\boldsymbol{\lambda}(\boldsymbol{z})$ is a one-to-one mapping of $\boldsymbol{\pi}(\boldsymbol{z})$ with no modelling restriction. However, as argued above, when $\boldsymbol{z}$ takes a large number of distinct configurations, to gather information from such sparse data, the different $\boldsymbol{\lambda}(\boldsymbol{z})$ may be constrained to satisfy a linear model

$$\boldsymbol{\pi}(\boldsymbol{z}) = g[\boldsymbol{\lambda}(\boldsymbol{z})] = g[\boldsymbol{Z}\boldsymbol{\psi}] \tag{3.1}$$

where the design matrix $\boldsymbol{Z}$ and the model parameters $\boldsymbol{\psi}$ are derived by stacking the following system of linear equations

$$\begin{aligned} \rho_i &= \alpha_i^X + \boldsymbol{z}_X^{'}\boldsymbol{\beta}_i^X,\ i = 1, \ldots, k-1 \\ \xi_j &= \alpha_j^Y + \boldsymbol{z}_Y^{'}\boldsymbol{\beta}_j^Y,\ j = 1, \ldots, k-1 \\ \tau_{ij} &= \alpha_{ij}^{XY} + \boldsymbol{z}_{XY}^{'}\boldsymbol{\beta}_{ij}^{XY},\ i, j = 1, \ldots, k-1 \end{aligned}$$

where $\boldsymbol{z}_X$, $\boldsymbol{z}_Y$ and $\boldsymbol{z}_{XY}$ denote respectively the subset of observed covariates $\boldsymbol{z}$ which are supposed to affect the marginal distribution of X and Y and their dependence structure.

3.5.1 Hypotheses of interest

A convenient feature of the parametrization defined above is that the hypothesis of stochastic monotonicity conditionally on relevant covariates can be expressed in the form of linear inequality constraints on the appropriate sub-vector of the $\boldsymbol{\psi}$. In particular, the hypothesis of stochastic monotonicity can be written as

$$\mathcal{H}_1 : \tau_{ij} = \alpha_{ij}^{XY} + \boldsymbol{z}_{XY}^{'}\boldsymbol{\beta}_{ij}^{XY} \geq 0\ \forall\ \boldsymbol{z}_{XY};\ i, j = 1, \ldots, k-1. \tag{3.2}$$

If we rewrite the set $\mathcal{H}_1$ compactly in terms of the model parameters as $\{\boldsymbol{\psi} : \boldsymbol{D}\boldsymbol{\psi} \geq \boldsymbol{0}\}$, notice that in typical applications the matrix $\boldsymbol{D}$ may have many more rows than columns and thus many inequalities are likely to be redundant. For example, in the application discussed below,there are more than 4000 inequalities with only 64 variables. There are algorithms, like the Fourier-Motzkin (see e.g. Schrijver (1986)) that could be used to spot and remove redundant inequalities; however, in our experience, redundant constraints do not slow significantly down the estimation algorithm anyway.

On the other hand, the hypothesis of equality of opportunities is equivalent to

$$\mathcal{H}_0 : \tau_{ij} = \alpha_{ij}^{XY} + \boldsymbol{z}_{XY}^{'}\boldsymbol{\beta}_{ij}^{XY} = 0 \; \forall \; \boldsymbol{z}_{XY}; \; i, j = 1, \ldots, k-1; \tag{3.3}$$

which can be equivalently rewritten in the standard form as

$$\mathcal{H}_0' : \boldsymbol{\alpha}^{XY} = \mathbf{0} \; \& \; \boldsymbol{\beta}_{ij}^{XY} = \mathbf{0}.$$

Finally, by $\mathcal{H}_2$ we will denote the unrestricted model.

3.5.2 Parameter estimates

Suppose now we have independent observations (X_i, Y_i, z_i) for a sample of n units. Let $\boldsymbol{t}(z_i)$ be a vector of size k^2 obtained by stacking one above the others the rows of a table having 1 in the cell X_i, Y_j and 0 elsewhere. To simplify notations, in the sequel we write $\boldsymbol{t}(i)$ instead of $\boldsymbol{t}(z_i)$; a similar convention will be adopted for any vector which depends on z_i. Under independent sampling, conditionally on z_i, $\boldsymbol{t}(i)$ has a multinomial distribution with vector of probabilities $\boldsymbol{\pi}(i)$. An algorithm for maximizing the multinomial log likelihood

$$L = \sum_i \boldsymbol{t}(i)' log[\boldsymbol{\pi}(i)] \tag{3.4}$$

is described by Colombi and Forcina (2001) and Dardanoni and Forcina (2008), and is based on an extension of an algorithm due to Aitchison and Silvey (1958). Essentially, at each step the algorithm does the following, until convergence:

- compute a quadratic approximation of the log likelihood in terms of the canonical (log-linear) parameters;
- compute a linear approximation of the canonical parameters in terms of $\boldsymbol{\psi}$,
- solve a weighted least square problem.

When inequality constraints are present, the weighted least square problem to be solved at each step requires a quadratic optimization which is itself iterative: there are many algorithms for quadratic optimization under inequality constraints, which are usually very fast and reliable.

3.5.3 Hypotheses testing

In the following let $\boldsymbol{\psi}_2$ denote the unrestricted maximum likelihood estimate (MLE) of $\boldsymbol{\psi}$, $\boldsymbol{\psi}_1$ be the MLE of $\boldsymbol{\psi}$ under the stochastic monotonicity hypothesis $\mathcal{H}_1$ and $\boldsymbol{\psi}_0$ be the MLE of $\boldsymbol{\psi}$ under the equality of opportunity hypothesis $\mathcal{H}_0$. Let $\boldsymbol{F}(\boldsymbol{\psi})$ denote the expected information matrix with respect to $\boldsymbol{\psi}$. From standard asymptotic results it follows that, if as n increases, $\boldsymbol{F}(\boldsymbol{\psi})/n$ is of full rank, $\boldsymbol{\psi}_2$ has an asymptotic normal distribution $N(\boldsymbol{\psi}, \boldsymbol{F}(\boldsymbol{\psi})^{-1})$. Therefore, hypotheses on single elements of $\boldsymbol{\psi}$ may be tested by comparing the estimate with the corresponding standard error. Joint testing may be based on the asymptotic distribution of the LR statistic. Recall the well known result that the LR for testing the unrestricted model against $\mathcal{H}_0$

$$T_{02} = 2(L(\boldsymbol{\psi}_2) - L(\boldsymbol{\psi}_0)) \tag{3.5}$$

has asymptotic χ^2_r distribution where r is the sum of the dimensions of $\boldsymbol{\alpha}^{XY}$ and $\boldsymbol{\beta}^{XY}$.

When inequalities are involved, the testing problem may be split into testing the unrestricted model $\mathcal{H}_2$ against $\mathcal{H}_1$ and testing $\mathcal{H}_1$ against $\mathcal{H}_0$. The corresponding LR statistics may be written as

$$T_{01} = 2(L(\boldsymbol{\psi}_1) - L(\boldsymbol{\psi}_0)) \tag{3.6}$$

$$T_{12} = 2(L(\boldsymbol{\psi}_2) - L(\boldsymbol{\psi}_1)) \tag{3.7}$$

It is also useful to recall the following:

Definition 1 *Let $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{V})$ be a k-dimensional normal random vector, and let C be a polyhedral cone in R^k. The squared norm of the projection of $\boldsymbol{b}$ onto C is a chi-bar-squared random variable $\bar{\chi}^2(C, \boldsymbol{V})$*

$$\bar{\chi}^2(C, \boldsymbol{V}) = \boldsymbol{b}'\boldsymbol{V}^{-1}\boldsymbol{b} - min_{\boldsymbol{a}\in C}(\boldsymbol{b} - \boldsymbol{a})'\boldsymbol{V}^{-1}(\boldsymbol{b} - \boldsymbol{a}) \tag{3.8}$$

and has distribution function:

$$Pr(\bar{\chi}^2(\mathcal{C}, \boldsymbol{V}) \leq x) = \sum_{i=0}^{k} w_i(\mathcal{C}, \boldsymbol{V}) F_\chi(x, i) \tag{3.9}$$

where $F_\chi(x, i)$ denotes the distribution function of a chi-square with i d.f. and $w_i(\mathcal{C}, \boldsymbol{V})$ is the probability that the projection of $\boldsymbol{b}$ onto $\mathcal{C}$ belongs to a face of dimension i.

We recall the following, which can be derived e.g. from Shapiro (1985) or Dardanoni and Forcina (1999):

Proposition 1 *Under the assumption that the true value ψ^o belongs to the interior of $\mathcal{H}_0$, the asymptotic distributions of T_{01} and T_{12} are:*

$$\begin{aligned} T_{01} &\rightarrow \bar{\chi}^2(\mathcal{H}_1, \boldsymbol{F}^{-1}(\psi^o)) \\ T_{12} &\rightarrow \bar{\chi}^2(\mathcal{H}_1^o, \boldsymbol{F}^{-1}(\psi^o)) \end{aligned} \tag{3.10}$$

where $\mathcal{H}_1^o$ is its dual of $\mathcal{H}_1$ in the metric determined by the information matrix at ψ^o.

Asymptotic critical values for these statistics depend on the probability weights $w_i(\mathcal{H}_1, \boldsymbol{F}^{-1}(\psi^o))$; unfortunately, except in very small problems, no closed form expression is available for the computation of these weights. However, reliable estimates may be obtained by Monte Carlo simulations as described by Dardanoni and Forcina (1998).

Il is worth recalling briefly the idea upon which the estimation of the probability weights is based. Let $\hat{\boldsymbol{b}}$ denote the projection of $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{V})$ onto $\mathcal{H}_1$, $\boldsymbol{D}(\boldsymbol{b})$ be the subset of rows of $\boldsymbol{D}$ such that $\boldsymbol{D}(\boldsymbol{b})\hat{\boldsymbol{b}} = \boldsymbol{0}$, and $\boldsymbol{Z}(\boldsymbol{b})$ the orthogonal complement of $\boldsymbol{D}(\boldsymbol{b})$. Then $w_g(\mathcal{H}_1, \boldsymbol{F}^{-1}(\psi^o))$ is the probability that $\boldsymbol{Z}(\boldsymbol{b})$ has rank g. Clearly, only non redundant rows can appear in $\boldsymbol{D}(\boldsymbol{b})$ and thus the presence of possibly redundant inequalities has no effect on the estimation of weights.

Since $\mathcal{H}_1$ is a composite hypothesis, one should search for the value of $\psi \in \mathcal{H}_1$ which gives the least favorable asymptotic null distribution for T_{12} and, as Wolak (1991) has shown, this value does not necessarily belongs to $\mathcal{H}_0$. Dardanoni and Forcina (1998) discuss some practical solutions to this problem. Finally, notice that the joint distribution of T_{01} and T_{12} can also be derived (see Dardanoni and Forcina (1999) for details), where use of this joint distribution for hypotheses testing is also compared with alternative testing procedures.

3.6 Testing conditional monotonicity: An application

In order to test for conditional monotonicity we use the National Child Development Study (NCDS), an ongoing survey that originally targeted over 17,000 babies born in Britain in the week 3-9 March 1958. Surviving members of this birth cohort have been surveyed on seven further occasions in order to monitor their changing health, education, social and economic circumstances: in 1965 (age 7), 1969 (age 11), 1974 (age 16), 1981 (age 23), 1991 (age 33), 1999 (age 41) and 2004 (age 46). At the age of 7, 11 and 16 mathematics, reading and general skills tests were taken by the cohort member, while at the age of 7 and 11 information on non-cognitive skills was also collected.

From the age of 16 individuals could leave education and enter the labor market. For those who stayed, the surveys from 1981 onwards together with a 1978 school survey provide information on the qualifications attained. Data on wages and social class were gathered at age 23, 33, 41 and 46. To study intergenerational mobility we also need data on parental socio-economic status. The first 4 surveys (1958,1965,1969,1974) contain data about parental background including age, education (1974), wage (1974), social class of father (1965, 1969, 1974) and mother (1974). These data sets therefore bring together information on socio-economic status for two consecutive generations.[4]

3.6.1 Choosing how to measure socio-economic status

To apply our stochastic monotonicity tests we first have to find suitable variables representing socio-economic status X and Y. Since true socio-economic status is not observed, intergenerational mobility scholars typically employ either wage (income) or social class in their analysis.

Economists often look at wages or income as the most important observed measure of socio-economic status. However both can be very sensitive to measurement error or temporary shocks such as short periods of unemployment, health shocks, or even short business cycles. In the

[4]The NCDS data managers have also collected information on the cohort members' children in 1991. However, back then these children were still very young and had not entered the labor market yet. No further information on these and new children of the cohort members was gathered in the 1999 and 2004 surveys.

standard linear model, using current wages rather than true lifetime socio-economic status can result in attenuation bias. Researchers try to solve this problem either by using average wage (income), whenever the data provide repeated observations, or by using an instrumental variable approach.[5] Notice that the attenuation bias holds also in our discrete mobility tables setting (Carroll, Ruppert, Stefanski, and Crainiceanu (2006) contains a thorough discussion of measurement error in non linear models); see e.g. Neuhaus (1999) for an analysis in the logit model.

On the other hand, sociologists prefer to use social class as a measure of lifetime socio-economic status (see for instance Erikson and Goldthorpe (2002)). They argue that not only social class is less sensitive to temporary shocks but also that it includes immaterial aspects such as prestige and power. The main limitations of social class originate from its subjectivity, since it is the researcher, using a combination of labor market occupation, education and other factors, that imputes the social class of the individual, sometimes also in an ordered manner, from the more prestigious occupation downwards. The way occupations are coded into social classes can sometimes affect the results. Moreover, the prestige associated with a social class can be time varying, i.e. being in, lets say, the skilled manual category in the 1960's is very much different than being in this category today, and this is a relevant problem in the case of intergenerational mobility where we look at individuals born twenty to forty years apart. Finally, within a class there could clearly be a large degree of heterogeneity; a painter and a carpenter may both be defined as skilled manual workers, but of course the socio-economic status of, say, Picasso is very different from that of an unknown painter. Yet, some of these problems affect wages (income) too. A miner might earn even more than an Academic professor due to the risk associated to his job, yet not many professors would choose to become miners.

Choosing how to measure socio-economic status inevitably depends on the data available. In our data there is not enough information to construct a reliable measure of father's permanent wage since this is observable only at one point in time. To overcome this problem Dearden, Machin, and Reed (1997) regress current wage on non time-varying factors such as education and social class, and then use the predicted variable as a measure of permanent wage. However, while there is no guarantee that this procedure eliminates attenuation bias, it also leads to a

[5]See Zimmerman (1992) for a discussion on the effect measurement error on measured mobility in the linear regression model.

mix of wage and social class mobility (because social class is used to predict wages) and it is not really suited to test for conditional mobility (because it directly uses education to predict wages). Moreover, as we show below, there are several individuals for whom wages are not available while we can observe social class. For these reasons, we consider social class a more reliable measure of lifetime socio-economic status than wages.

In this application we use the 1991 data on the cohort member socio-economic status coupled with the 1974 data on father's status. These are also the NCDS surveys used by Dearden et al. (1997) and Blanden et al. (2007) in their studies on wage mobility. We first select all male cohort members for which we observe cognitive and non-cognitive skills at age 7, 11 and 16 (cognitive skills only), educational attainment and father's age. We then select those individual's for whom we observe both social class in 1991 and father's class in 1974.[6]

Table 3.1 shows summary statistics for the social class measures. Social Class is a status variable grouping occupations into six broad categories, ordered on a skill basis.[7] In the data, sons are more skilled than their parents were. The distribution of son's socio-economic status actually stochastically dominates the father's one.

Table 3.1: Social Class - Males

	Son		Father	
Professional	7.21	(7.21)	5.97	(5.97)
Intermediate	35.84	(43.05)	21.63	(27.60)
Skilled Non-manual	12.92	(55.97)	10.87	(38.47)
Skilled Manual	28.89	(84.86)	45.01	(83.47)
Semiskilled	12.77	(97.63)	13.18	(96.65)
Unskilled	2.37	(100.00)	3.35	(100.00)
Observations	1942		1942	

Values are in percentages. Numbers in brackets are cumulated percentages.

Table 3.2 shows summary statistics for parent's and son's weekly net wages, son's highest educational qualification and father's age. The NCDS collected information on parental net wages only in 1974, with separate questions about father's and mother's wages and other sources of income. Note that in the original coding wage was grouped into 12 wage bands and we assign

[6]We select males only to make our results comparable to previous studies.

[7]Social Class variables are derived according to the Registrar Classification (RG). This classification imputes social class using only information about occupation. This is a quite common and simple grouping methodology, even though some sociologists have proposed alternative ones. The Goldthorpe class schema for instance (see Erikson and Goldthorpe (2002)) aims to capture qualitative differences in employment relations. Unfortunately, the classes distinguished by this schema are not consistently ordered according to some inherent hierarchical principle. Therefore the Goldhtorpe class schema does not suit our statistical model.

to each observation the median value of the observed band. Finally, in the data there are several individuals for whom father's social class are available but wages are not, while the opposite is quite rare. There are a number of explanations. Some individuals (or their fathers) are self-employed, and their wages are not reliable. For other individuals the wage is not available either because they were unemployed or because they chose not to report it.

Table 3.2: Summary Statistics

	Obs	Mean	S.D.
Son's Net Wage	1341	301.07	102.79
Father's Net Wage	1486	232.56	83.93
No Qualification	1942	0.45	0.49
O Levels	1942	0.33	0.46
A Levels	1942	0.09	0.29
Higher Education	1942	0.13	0.33
Father's Age	1942	46.55	6.12

Father and Son net wages are in January 2001 prices.

3.6.2 Mobility Tables

In this section we present some unconditional transition matrices. We group social classes into three categories roughly corresponding to a high/medium/low skilled partition.

3 Professional+Intermediate

2 Skilled Non-Manual + Skilled Manual

1 Semiskilled + Unskilled

There is no compelling reason to use three categories rather than two, four or any other number. Having more categories allows the researcher to have a better understanding of the heterogeneity across the joint distribution but at the same time it can make the identification of the parameters cumbersome. The larger the number of categories the lower the sample size within each cell. This problem only exacerbates when conditioning on other covariates. Given the sample size of our data, the likelihood function failed to converge when using more that three categories and our preferred set of control variables.

Table 3.3 shows the mobility tables using son's social class in 1991 and father's class in 1974. In the table we also include the Person's chi-square statistic for independence, with the degrees of freedom in brackets. The chi-square statistic is very large leading to reject the hypothesis of independence.

Table 3.3: Social Class Mobility Table

Father/Son Class	1	2	3	
1	4.53	8.08	3.91	16.53
2	8.81	26.06	21.01	55.87
3	1.80	7.67	18.13	27.60
	15.14	41.81	43.05	100.00

1942 observations. Numbers on table are percentages.
Chi-Square (4) = 192.73

Table 3.4 shows the *Local-Global Log-Odds Ratios* in the unconditional table. The first two odds correspond to the first two rows of the mobility table. If both τ_{11} and τ_{12} are positive than the first row (low class) is stochastically dominated by the second one (medium class). The same reasoning applied to τ_{21} and τ_{22} (medium class is stochastically dominated by high class). All the odds in the table are positive. Yet the odds $[\tau_{11},\tau_{12}]$ are smaller than the $[\tau_{21},\tau_{22}]$ indicating stronger stochastic dominance towards the top of the joint distribution.

Table 3.4: Local Global Odds Ratios

Row	Log-Odds	
R_{12}	τ_{11}	0.7181
	τ_{12}	0.6641
R_{23}	τ_{21}	0.9851
	τ_{22}	1.1551

3.6.3 Control Variables

As we explained in the section 3.5, our aim is to test for dependence conditional on some characteristics of the parents and offsprings. Since most economic models of intergenerational mobility assume that the transmission of the ability endowment across generations is one of the main reasons behind immobility (see Becker and Tomes (1979) or Grawe and Mulligan (2002) for instance) a starting point is to investigate monotonicity conditional on cognitive skills. However, as Becker and Tomes (1979), Bowles and Gintis (2002), Erikson and Goldthorpe (2002), Dardanoni et al. (2006) suggest, cognitive ability is just one dimension of the endowment stock. Recently Heckman et al. (2006) and Cunha and Heckman (2007) show that non-cognitive skills can also explain a diverse array of outcomes such as schooling choices, wages, employment and work experience. It is quite likely that non-cognitive skills are also transmitted across generations, if not genetically, because of parental behavior and education. Finally, the human capital models predicts that high-status parents invest more in their children. In turn this implies that these children have more human capital. Therefore we choose to test for monotonicity conditional on the educational attainment, cognitive and non-cognitive skills of the offspring (son). Since the fathers were of different ages at the moment of the survey, we also control for father's age.

Given the education system faced by the 1958 cohort, its educational attainment is measured by 4 ordered categories corresponding to 'No Qualification', 'O Levels', 'A Levels' and 'Higher Education'.[8] To control for cognitive skills we use the mathematics and reading test scores. These tests were taken by the cohort members at the age of 7, 11 and 16. We use all these multiple age-skills observations but in order to reduce the control variables space, at each age we replace the original maths and reading scores with the principal component. In all cases the principal component explains no less than 90% of the total variance. In the case of non-cognitive skills things are a bit more complex. Both at age 7 and age 11 there are 12 scores for non-cognitive skills, such as depression, anxiety, hostility etc., as reported by teachers in schools. (No score is available at age 16.) In order to keep our problem computationally tractable, we do a factor analysis of the non-cognitive scores using the iterated principal factor method. Out of 12 scores, only two eigenvalues are larger than 1, with the third being sensibly smaller. Therefore, at each age point, we retain only two factors. In Table 3.5 we show the rotated loading factors. The first factor captures the skill to relate to other individuals, either adults or other children. The second factor captures emotional problems. There are no large differences between age 7 and 11. The final factors are obtained using the regression method. [9]

Finally, since father's age was recorded only in the original 1958 survey, we restrict our sample

[8]In the UK schooling is compulsory up to the age of 16, when individuals can, at the end of the scholastic year, stay in education or enter the labor market. Those who stay on at age 16 enrol for the O Levels or CSE qualifications, which are taken immediately at the end of the scholastic year. These students are still aged 16 when they obtain the qualification. In the Autumn term of the same year, those who successfully obtained 5 or more O Levels/CSE can enrol for A Levels. These last 2 years, until individuals are aged 18. Passing 2 A-level constitutes the minimum level required for entry in Higher Education. Once the student has completed A Level, he can gain admission to the Universities, Polytechnic or Colleges of Higher Education where a first degree is obtained. The time needed to gain a degree varies by subject but in the majority of cases it takes 3 years.

[9]Note that for the factor analysis we use a much larger sample since we do not have to take into account missing values in social class or other covariates.

Table 3.5: Loading Factors - Non-cognitive scores

	Age 7			Age 11		
	F1	F2	Uniquens.	F1	F2	Uniquens.
Unforthcomingness	-0.0690	0.7279	0.4654	-0.0807	0.7161	0.4807
Withdrawal	0.0649	0.6693	0.5478	0.0908	0.6856	0.5217
Depression	0.2727	0.6757	0.4691	0.3496	0.6305	0.4802
Anxiety for acceptance of adults	0.4061	-0.1096	0.8230	0.4257	-0.0487	0.8164
Hostility towards adults	0.6042	0.2118	0.5901	0.6455	0.1988	0.5438
Writing off adults and standards	0.4362	0.5162	0.5433	0.4696	0.4690	0.5595
Anxiety for acceptance by kids	0.6662	-0.0429	0.5543	0.6862	-0.0386	0.5276
Hostility towards children	0.6736	0.0942	0.5373	0.6593	0.1311	0.5481
Restlessness	0.5467	0.1447	0.6802	0.5258	0.1190	0.7093
Inconsequential behavior	0.7761	0.2020	0.3569	0.7771	0.1931	0.3588
Miscellaneous symptoms	0.2354	0.5894	0.5972	0.3329	0.5408	0.5967
Miscellaneous nervous symptoms	0.2636	0.1910	0.8941	0.3057	0.1689	0.8780

Number of obs = 14931 (Age 7), 14158 (Age 11). Retained factors = 2. Number of params = 23.

to those cohort members living with a biological father during their childhood.

3.6.4 Equality of Opportunity

Table 3.6 presents results when we test for equality of opportunity (independence) as shown in section 3.5.1 (equation 3.3) and section 3.5.3 (equation 3.5). To begin with we do not condition on any covariates, and test for equality of opportunity by restricting the odds to be zero. This is done by setting the constant terms α^{XY} equal to zero. Since we are ultimately interested in stochastic dominance, that is whether one row "dominates" the other, we first impose the restrictions by looking at pairs of rows. Therefore, we first impose $[\tau_{11}, \tau_{12}] = 0$ (R_{12}) and then $[\tau_{21}, \tau_{22}] = 0$ (R_{23}) . We then impose the restriction on all rows simultaneously: $\boldsymbol{\tau} = 0$ (AR). The numbers in the table are the p-values (likelihood-ratios in brackets) from a standard chi-square distribution. The numbers in bold indicate the number of restrictions. In the unconditional case there are only the constant terms to restrict. All the p-values are equal to zero at the fourth decimal digit. When imposing the equality of opportunity restriction on all the rows the likelihood ratio is very close to the chi-square statistic in table 3.3 as expected.

Next, we test independence conditional on our set of covariates: cognitive skills (measured at age 7, 11 and 16), non-cognitive skills (measured at age 7 and 11) up to the third principal component, educational attainment (at age 23) and father's age. There are 13 variables in total. Table 3.12 in the Appendix presents the estimated coefficients and standard errors for the unrestricted model.

Table 3.6: P-values Independence

Row	Unconditional		Conditional	
	P-values	Restrict.	P-values	Restrict.
R_{12}	0.0000	**2**	0.0160	**24**
	(31.3557)		(41.1642)	
R_{23}	0.0000	**2**	0.0007	**24**
	(114.3083)		(52.4605)	
AR	0.0000	**4**	0.0000	**48**
	(189.9833)		(103.1234)	

Likelihood ratios in brackets.

We model the four odds ratios ($\boldsymbol{\tau}$) and two marginal parameters for the son ($\boldsymbol{\xi}$) as a function of the full set of covariates. On the other hand we model the two marginal parameters for the father ($\boldsymbol{\rho}$) as a function of father's age only, that is we rule out that father's class can be a function of son's education, cognitive and non-cognitive skills. Note also that since all the covariates are centered, the constant terms have a direct interpretation. They measure the *Local-Global Log-Odds Ratios* for the average individual. For $\boldsymbol{\xi}$, education and cognitive skills coefficients have the expected positive sign, meaning that those sons with better education and cognitive skills are more likely to have a higher status. None of the non-cognitive skills factors is statistically significant on its own. When interpreting the coefficients for the four odds ratios ($\boldsymbol{\tau}$), one has to remember that a larger τ indicates less mobility. For instance, the Higher Education coefficient is positive and statistically significant for τ_{11} and τ_{12}. Intuitively, this result suggests that those individuals with Higher Education in the medium class were very unlikely to have a father in low class, and at the same time those individuals with Higher Education in the low class were unlikely to have a father in medium class.

In last two columns of Table 3.6 we present the conditional p-values. To test for independence we impose 12 restrictions given by 1 constant term + 11 coefficients (one for each covariate) for each LG log-odds ratio. Thus the total number of restrictions equal 12 times the number of odds. Once we condition on this rich set of covariates the p-values become larger. At a 1% significance level independence is no longer rejected in the first row. This result confirms that our set of covariates do explain part of the mobility mechanism.

3.6.5 Conditional stochastic monotonicity

We now formally test our mobility tables for equality of opportunity against stochastic monotonicity, as explained in section 3.5.3, equation (3.6). To test the null hypothesis we need to compute the $\bar{\chi}^2$ distribution function. Being one-directional, this test is more powerful than the previous one where the null hypothesis of equality of opportunity (independence) is tested against the unrestricted model. Table 3.7 column (*EoP vs SM*) shows the computed *p*-values obtained under the null hypothesis of equality of opportunity against stochastic monotonicity conditional on our rich set of control variables. As expected, since this test is more powerful, all the conditional *p*-values are now smaller than in table 3.6. We now reject independence even at a 1% significance level.

Since this results suggests that independence is rejected even conditionally on a rich set of controls, we now test whether our data give evidence of conditional stochastic monotonicity. In column (*SM vs Unr*) we display the *p*-values obtained under the null hypothesis of stochastic monotonicity against the unrestricted model as shown in section 3.5.3, equation (3.7). The *p*-values are all very high meaning that there is strong evidence of stochastic monotonicity, or in other words very few *Local-Global Log-Odds Ratios* are negative in the sub-tables when conditioning on z.

Table 3.7: P-values Stochastic Monotonicity

Row	*EoP vs SM*	*SM vs Unr*
R_{12}	0.0003	0.8450
R_{23}	0.0000	0.8284
AR	0.0000	0.9460

Numbers in table are *p*-values.

3.6.6 Wage Mobility

For completeness, in this section we replicate the analysis using the wage mobility table illustrated in table 3.8. As we discussed above, the sample size is much smaller. As in the case of social class, we group the individuals into three categories based on their wage percentile. Unfortunately, given that the original father's wage variable was coded into 12 bands, it is not possible to exactly partition it in three terciles. The chi-square statistic is smaller than for social class (table 3.3) though we still reject independence.

Table 3.8: Wage Class Mobility Table

Father/Son Class	1	2	3	
1	14.04	10.78	8.79	33.61
2	12.41	10.51	10.42	33.33
3	7.61	11.32	14.13	33.06
	34.06	32.61	33.33	100.00

1104 observations. Chi-Square (4) = 37.03.

In table 3.9 we test for equality of opportunity (independence) both unconditional and conditional on our rich set of control variables. In the Appendix, table 3.13, we show the estimated coefficients and standard errors for the unrestricted model. Anew, the p-values become larger once we condition on z. Note however, that for R_{12} we can not reject unconditional independence. Overall the p-values are larger than in the case of social class. From a comparison of tables 3.12 and 3.13 we also notice that three out of four *Local-Global Log-Odds Ratios* constant terms (the odds for the average individual) are larger when using social class. This difference between wage and social class might be due to measurement error or temporary shocks affecting wages more than social class. However, we can not rule out that sample selection might also be driving this difference. While a full comparison of social versus wage class is beyond the scope of the paper, we tried to investigate this issue further by re-estimating the odds using only those son/father pairs for whom we observe both social and wage class. Unfortunately, given the further restriction in the number of observations and the still large number of covariates, the likelihood failed to converge.

Table 3.9: P-values Independence - Wage Mobility

Row	Unconditional		Conditional	
	P-values	Restrict.	P-values	Restrict.
R_{12}	0.3267	**2**	0.4808	**24**
	(2.2372)		(23.6668)	
R_{23}	0.0001	**2**	0.0002	**24**
	(18.8047)		(55.8950)	
AR	0.0000	**4**	0.0044	**48**
	(36.0567)		(77.5813)	

Likelihood ratios in brackets.

In table 3.10 we show the p-values obtained while testing for stochastic monotonicity. As for social class, when testing *EoP* against stochastic monotonicity, column (*EoP vs SM*), all the conditional p-values are now smaller than in table 3.9, though we still can not reject independence in the first row (R_{12}). Column (*SM vs Unr*) provide again evidence of conditional stochastic monotonicity.

Table 3.10: P-values Stochastic Monotonicity - Wage Mobility

Row	*EoP vs SM*	*SM vs Unr*
R_{12}	0.3342	0.5574
R_{23}	0.0005	0.0701
AR	0.0007	0.3064

Numbers in table are p-values.

3.7 Conclusions

The aim of this paper was to test for stochastic monotonicity in intergenerational socio-economic mobility tables. To do so we apply and extend the methodology discussed in Dardanoni and Forcina (1998) and Bartolucci et al. (2001). We first test for unconditional stochastic monotonicity using a set of 149 intergenerational mobility tables in 35 different countries, where it emerges that monotonicity cannot be rejected in hardly any table.

We then explain how a number of controls such as education, cognitive and non-cognitive skills can be used to investigate whether monotonicity still holds after conditioning on these factors. One main contribution of this paper is to formalize a test of dependence in mobility tables, using both continuous and discrete control variables. In the economic literature, no previous work on intergenerational mobility tables has dealt with continuous controls. Since current research on mobility is focussing on the determinants dependence in socio-economic status between parents and offspring, conditioning on discrete and continuous covariates is increasingly important. We consider our methodology an important tool for future research, with potential applications to other empirical fields besides intergenerational mobility.

To apply our test of conditional monotonicity we use the NCDS, a UK cohort data with information on the socio-economic status of the cohort members and their parents, individual's educational qualifications, cognitive and non-cognitive skills. Our tests show evidence of stochastic monotonicity both unconditionally and conditionally. While it is not surprising that the unconditional joint distribution exhibits monotonicity, it is interesting to find that such a strong form of dependence subsists even conditional on educational achievement, cognitive and non-cognitive skills. This result reinforces the findings of Solon (1999), Bowles and Gintis (2002), Restuccia and Urrutia (2004), Dardanoni et al. (2006), Blanden et al. (2007) indicating that part of the mechanism linking parent's and offspring's socio-economic status is still a black box.

Finally, we observe only minor differences between social and wage class tables, where if anything there seems to be more dependence when using social rather than wage class.

Table 3.11: Ganzeboom, Luijks and Treiman tables

Country/yr	LR	constr.	Country/yr	LR	constr.	Country/yr	LR	constr.
AUS65	16.6230	6	GER80a	3.3808	4	NOR72l	5.6674	8
AUS67	4.5578	5	GER80p	4.1145	9	NOR72s	3.0935	3
AUS67l	5.1737	8	GER80z	4.5121	7	NOR73	8.9025	4
AUS73	11.239	4	GER82a	12.5920	6	NOR82w	4.2696	8
AUS87	3.8706	6	GER84a	1.6064	4	NZE76	2.8615	4
AUT69n	5.2945	4	HKG67	1.8590	8	PHI68	66.9629	1
AUT74p	5.0644	5	HUN62	70.2130	2	PHI73	9.2018	2
AUT78	15.4645	4	HUN73	28.5077	4	POL72	50.5442	4
BEL71e	13.7803	6	HUN73l	16.5144	4	POL82	3.5511	5
BEL75	11.0925	5	HUN82	1.3950	3	POL87	9.7428	6
BEL76	5.4211	4	HUN83	26.3894	3	PUE54	0.1228	1
BRA73	9.4740	3	HUN86	5.3734	6	QUE60	9.7055	8
CAN73	4.4114	4	IND62c	8.3105	6	QUE73	4.9998	3
CAN82w	3.9487	5	IND63a	14.2253	8	QUE77	1.4496	6
CAN84	3.4051	3	IND63c	26.7984	7	SCO74	8.9427	6
CSK67	21.2481	2	IND71n	5.2398	4	SCO75	16.0839	5
DEN71	6.4718	9	IRE74	8.1673	3	SPA65	2.8188	2
DEN72l	7.5135	6	ISR62c	8.9032	7	SPA67t	9.1810	7
DEN72s	1.7511	4	ISR74	4.6268	1	SPA75	68.8972	4
ENG51	5.0602	4	ITA63	6.5001	4	SWE50	9.1278	5
ENG63	1.2945	5	ITA68	4.3434	6	SWE60	1.9379	4
ENG67t	0.2951	6	ITA72	9.5852	6	SWE72l	5.5451	7
ENG69	0.9969	4	ITA74	6.6980	4	SWE72s	2.0677	3
ENG72	10.4063	4	ITA75p	14.7003	7	SWE73	0.2150	2
ENG74	2.7902	5	JAP55	6.6026	4	SWE83w	4.7170	5
ENG74p	6.0480	5	JAP65	1.0588	5	SWI76p	4.1620	4
ENG83	3.4236	3	JAP67	7.6923	7	TAI70	10.0489	9
ENG86	2.0181	5	JAP69t	4.3844	9	TAI70l	7.6641	6
FIN67t	8.6317	10	JAP71n	2.2285	3	USA47	4.5285	4
FIN72l	6.8491	8	JAP75	1.0965	5	USA47l	1.7094	5
FIN72s	3.5080	8	MAL67	22.1726	7	USA59c	3.2706	5
FIN75p	2.1559	6	NET58	6.3140	6	USA62o	2.3330	3
FIN80	12.2139	4	NET67t	2.6724	7	USA72g	5.7139	8
FIN82w	10.7844	9	NET70	2.6992	4	USA73g	8.7437	8
FRA58	4.6786	4	NET71	2.4754	5	USA73o	7.7145	3
FRA64	9.3367	4	NET71e	3.6097	5	USA74g	3.3402	5
FRA67	7.8008	8	NET74p	1.1753	4	USA74p	8.4365	5
FRA70	11.7893	2	NET76	1.2016	4	USA75g	2.8330	7
FRA71e	7.0699	4	NET77	5.8811	4	USA76g	3.3676	5
GER59	3.0701	5	NET77x	2.1617	3	USA77g	5.7246	4
GER69	3.3086	8	NET79p	12.1392	8	USA78g	4.8763	6
GER69k	10.5659	5	NET82	4.6030	7	USA80g	5.0530	7
GER75p	4.4880	5	NET82u	5.1157	8	USA81w	1.8263	5
GER76z	8.3095	6	NET85	1.5353	4	USA82g	7.8830	5
GER77z	3.6863	6	NIG71n	9.6726	5	USA83g	3.0734	4
GER78	6.0179	7	NIR68	2.5272	7	USA84g	4.5427	8
GER78x	3.5050	7	NIR73	4.1850	6	USA85g	6.5725	5
GER78z	6.2942	5	NOR57	11.0213	7	USA86g	6.6470	8
GER79z	2.2963	8	NOR65	1.1214	4	YUG67t	6.2877	8
GER80	4.8059	5	NOR67t	1.8285	6	—	—	—

Table 3.12: Unrestricted Model - Social Class (1974)

Parameter	Marginal							
	ρ_1		ρ_2		ξ_1		ξ_2	
Constant	1.6352	(26.4510)	-0.9620	(-18.9698)	2.1078	(24.1718)	-0.2674	(-5.2418)
Father Age	-0.0338	(-3.5990)	-0.0003	(-0.0396)	-0.0290	(-2.8677)	-0.0004	(-0.0438)
O Level	—	(—)	—	(—)	0.5343	(2.9851)	0.2329	(1.8124)
A Level	—	(—)	—	(—)	1.6015	(3.0112)	1.0127	(4.7179)
High. Educ.	—	(—)	—	(—)	0.7777	(2.2970)	1.3276	(6.5289)
C. Skills 7	—	(—)	—	(—)	-0.0798	(-1.0457)	-0.0537	(-0.8778)
C. Skills 11	—	(—)	—	(—)	0.1519	(1.4253)	-0.0109	(-0.1395)
C. Skills 16	—	(—)	—	(—)	0.4170	(3.9301)	0.5326	(6.1489)
NC. Skills 7 (1st)	—	(—)	—	(—)	0.0127	(0.1694)	0.0089	(0.1438)
NC. Skills 7 (2nd)	—	(—)	—	(—)	-0.0371	(-0.4785)	-0.0749	(-1.0838)
NC. Skills 11 (1st)	—	(—)	—	(—)	0.0309	(0.4194)	-0.0038	(-0.0592)
NC. Skills 11 (2nd)	—	(—)	—	(—)	-0.1022	(-1.4225)	-0.0955	(-1.4514)
Parameter	**Odds**							
	τ_{11}		τ_{12}		τ_{21}		τ_{22}	
Constant	0.7785	(3.4852)	0.5249	(3.0082)	0.4118	(1.5316)	0.6935	(5.7366)
Father Age	0.0255	(0.9328)	0.0517	(2.0670)	-0.0878	(-3.1548)	-0.0375	(-1.9414)
O Level	-0.4090	(-0.8632)	-0.0218	(-0.0527)	0.4563	(0.9364)	0.1094	(0.3608)
A Level	0.9018	(0.6769)	0.0666	(0.1026)	-0.1924	(-0.1311)	-0.1019	(-0.1968)
High. Educ.	1.7188	(1.9630)	1.7835	(2.9363)	0.8091	(0.6344)	-0.1658	(-0.3114)
C. Skills 7	0.1127	(0.5435)	0.0896	(0.4560)	-0.2430	(-1.1552)	0.0658	(0.4555)
C. Skills 11	-0.0001	(-0.0004)	0.2169	(0.9342)	-0.2386	(-0.8128)	-0.1569	(-0.8348)
C. Skills 16	-0.0305	(-0.1055)	-0.8564	(-2.9761)	-0.0636	(-0.2181)	-0.1141	(-0.5569)
NC. Skills 7 (1st)	0.1527	(0.7321)	-0.0168	(-0.0837)	-0.4423	(-2.1842)	-0.1249	(-0.8676)
NC. Skills 7 (2nd)	-0.2082	(-0.8959)	-0.0596	(-0.2756)	-0.4707	(-2.2672)	-0.2638	(-1.6378)
NC. Skills 11 (1st)	-0.1114	(-0.5616)	0.2076	(0.9010)	0.5944	(2.4360)	-0.0147	(-0.0987)
NC. Skills 11 (2nd)	-0.0560	(-0.2812)	-0.1678	(-0.8019)	0.0958	(0.5054)	0.0340	(0.2208)

Columns 1 and 2 correspond to the father's marginals.

Table 3.13: Unrestricted Model - Wage Class (1974)

Parameter	Marginal							
	ρ_1		ρ_2		ξ_1		ξ_2	
Constant	0.7025	(10.8281)	-0.7153	(-11.0466)	0.7790	(11.3201)	-0.7829	(-11.2551)
Father Age	-0.0604	(-5.3298)	-0.0491	(-4.1184)	0.0103	(0.8803)	-0.0031	(-0.2556)
O Level	—	(—)	—	(—)	0.1923	(1.1480)	0.0386	(0.2096)
A Level	—	(—)	—	(—)	0.6051	(2.0390)	0.5021	(1.9052)
High. Educ.	—	(—)	—	(—)	0.7753	(2.7621)	0.7313	(2.9104)
C. Skills 7	—	(—)	—	(—)	0.0499	(0.6402)	0.1655	(1.9331)
C. Skills 11	—	(—)	—	(—)	0.0086	(0.0813)	0.0760	(0.7121)
C. Skills 16	—	(—)	—	(—)	0.3645	(3.2359)	0.2502	(2.0866)
NC. Skills 7 (1st)	—	(—)	—	(—)	-0.0552	(-0.6698)	-0.1992	(-2.0448)
NC. Skills 7 (2nd)	—	(—)	—	(—)	0.0268	(0.3110)	0.1316	(1.4177)
NC. Skills 11 (1st)	—	(—)	—	(—)	0.0434	(0.5449)	0.1690	(1.9518)
NC. Skills 11 (2nd)	—	(—)	—	(—)	-0.2199	(-2.7225)	-0.3527	(-3.5865)
Parameter	Odds							
	τ_{11}		τ_{12}		τ_{21}		τ_{22}	
Constant	0.1589	(0.8712)	0.1128	(0.5894)	0.5186	(2.6211)	0.3566	(1.9529)
Father Age	-0.0529	(-1.7568)	-0.0502	(-1.5437)	0.0973	(2.8043)	0.0001	(0.0024)
O Level	0.4536	(1.0696)	0.6547	(1.3205)	0.3801	(0.8438)	-0.4857	(-1.0260)
A Level	2.3471	(2.4014)	1.7889	(2.5130)	-2.8069	(-2.8402)	-2.1742	(-3.0755)
High. Educ.	0.6454	(0.9285)	0.6428	(0.9749)	-0.2310	(-0.3092)	-0.3686	(-0.5742)
C. Skills 7	0.4129	(2.0222)	0.2796	(1.1973)	-0.1209	(-0.5801)	-0.3467	(-1.5591)
C. Skills 11	-0.5185	(-1.8872)	-0.2244	(-0.7904)	0.6906	(2.3944)	0.1713	(0.6224)
C. Skills 16	0.1640	(0.5558)	0.1466	(0.4545)	-0.5773	(-1.8904)	-0.2207	(-0.7104)
NC. Skills 7 (1st)	0.0837	(0.3861)	0.1728	(0.6398)	-0.0475	(-0.2135)	-0.2570	(-1.0288)
NC. Skills 7 (2nd)	0.2386	(1.0826)	0.0087	(0.0363)	-0.0420	(-0.1831)	-0.4537	(-1.8804)
NC. Skills 11 (1st)	0.1253	(0.5918)	0.4016	(1.5518)	0.4297	(1.7630)	0.0807	(0.3501)
NC. Skills 11 (2nd)	-0.0871	(-0.4103)	0.1781	(0.6943)	0.1454	(0.6697)	-0.3068	(-1.1944)

Columns 1 and 2 correspond to the father's marginals.

CHAPTER 4

THE EFFECT OF HOME COMPUTER USE ON CHILDREN'S COGNITIVE AND NON-COGNITIVE SKILLS

4.1 Introduction

In the last decade a number of papers have stressed that educational and labor market outcomes are largely pre-determined by the cognitive and non-cognitive skills accumulated during early childhood. Keane and Wolpin (1997, 2001), Cameron and Heckman (1998, 2001) have found that in the US individual educational decisions are mainly driven by cognitive skills such as maths and verbal skills. Those with good skill endowments by age 16 are much more likely to enrol and complete college education. Financial constraints are either not binding or most individuals manage to offset them by working part-time and borrowing. Their results suggest that policies targeting educational attainment or educational disparities between Black, Hispanic and White youth must act on these skill inputs to be effective. Heckman et al. (2006) find that a low-dimensional model of cognitive and non-cognitive abilities explains a diverse array of outcomes such as schooling choices, wages, employment, work experience, choice of occupation but also a variety of adolescent risky behaviors such as criminality, cigarette smoking and alcohol use. Cunha et al. (2006) review the evidence on the life cycle of human skill formation. They conclude that ability gaps in both cognitive and non-cognitive skills across individuals and across socioeconomic groups open up early in the life cycle and IQ deficits need to be addressed at very early ages for interventions to be effective.

Given this evidence, there is a growing interest in estimating the skills production function. Researchers are trying to uncover the main inputs and their time varying effect (see Todd and Wolpin (2003) and Cunha and Heckman (2007) for a discussion). However estimating the causal effect of these inputs is difficult because all sorts of endogeneity problems might lead to inconsistent estimates and economists have mainly focused on a few inputs that are either very important or for which experimental designs are available. To mention only a few recent studies that have looked at the determinants of math and reading achievements, Rivkin, Hanushek, and Kain (2005) analyze the effect of teacher quality, Dahl and Lochner (2005) and Belley and Lochner (2007) estimate the effect of parental income, Bernal and Keane (2008) and Berlinski et al. (2008) evaluate the effect of respectively child care and pre-school while Gentzkow and Shapiro (2008) identify the effect of pre-school television exposure.

The objective of this paper is to investigate the effect of using a computer at home. Computers are a relatively new input in the production function. Figure 4.1 shows OECD data on home computer access in a few selected countries. [1] There is a clear upward trend. Since 2005, in all the countries 70% or more of the households had a computer at home and this proportion is likely to rise further. As we show below, children use home computers quite extensively. Yet, little is known about the effect of computer usage on children's cognitive and non-cognitive skills.

Psychologists have long investigated the effect of time spent in front of the television on children's development, see Schmidt and Anderson (2007) for a review, and are now shifting their attention to computers, see Subrahmanyam, Greenfield, Kraut, and Gross (2001). Even though computers and TV are different media devices, understanding why TV time can have an effect on children's skills is a useful starting point to analyze the effect of computer time. There are three main theories in psychology. The first theory emphasizes the effect of TV content, that is what matters is what children watch and not TV time per se. On the one hand, this theory states that educational programs can have a positive effect on skills. On the other hand, if children watch mostly cartoons or general entertainment programs, TV would have no impact. The second theory points at the time allocation problem. Children, like adults, have a limited time endowment. The more time is spent watching TV, the less time is available for other activities. If TV time displaces other educational or social activities then it might have an effect even irrespectively of what children watch. The third theory points at the passive nature

[1] Data for the USA is available for only a few years.

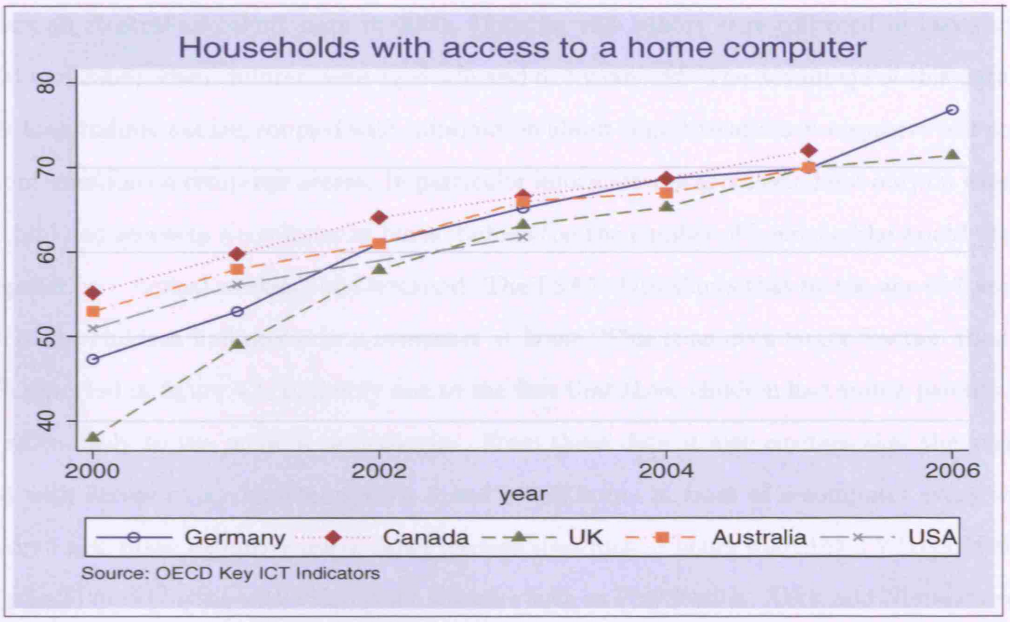

Figure 4.1: Household Home Computer Access

of television. Viewing requires little overt behavior, programs are visually explicit and require little visual imagination, and the medium is not interactive. As a result children might become intellectually passive. While we refer the reader to section 4.2 for a more complete review of the literature, we anticipate that psychology studies conclude that the effect of watching TV strongly depends on the content of the programs watched (educational programs having a positive effect) and on the socio-economic status of the parents (children with low status parents benefiting more from TV), the latter reflecting the quality of those activities displaced by TV time. Computers and TV share some similarities but there are also major differences. Computers imply more freedom with respect to content, since there is a very large variety of software or internet content to choose from. Computers are also more interactive than TV, with most software requiring continuous inputs from the users. Still, most of the above discussion can be extended to computer time. Content can matter, other activities will be displaced by computer time, and computer use can be intellectually challenging (rather than intellectually passive).

In this paper we use the Longitudinal Study of Australian Children (LSAC) data, which follows an Australian cohort born in 2000. Data for this cohort were collected in two surveys (2004 and 2006) when children were aged 4/5 and 6/7 years old. The advantage of this data lies in its longitudinal nature, coupled with information about cognitive and non-cognitive test scores and information on computer access. In particular information was collected not only on whether the child had access to a computer at home, but also on the number of hours he/she would use the computer on a typical weekday and weekend. The LSAC data shows that by the age of 7 around 88% of the children had access to a computer at home. This is an even larger fraction than the 70 % reported in figure 4.1, probably due to the fact that these children had young parents who are more likely to use modern technologies. From these data it also emerges that the average child with access to a computer spends 3 and a half hours in front of a computer every week. Children also make extensive use of other devices spending 13 hours watching TV/DVD's and 3 and a half hours playing with video game consoles such as PlayStation, XBox and Nintendo every week. If we are interested in the skill production function we can not neglect the importance of these inputs given that they absorb a considerable amount of time. Here we mainly focus on computer use though we also try to shed some light on the effect of TV/DVD's and video game use. We look at both cognitive and non-cognitive skills. Both types of skills might be affected by the content (educational software, games, emailing or messaging, other internet use) but also by the displaced activities. If, for instance, computer time displaces reading books or time spent on homework, cognitive skills might be affected. Similarly, if computer time displaces social activities, with parents or other children, non-cognitive skills could be influenced.

Previous research has focused on the effect of TV on skills, of computers in schools or on the effect of a home computer on high school completion. Our paper contributes to the existing human capital literature by focusing on the effect of home computer use on early childhood cognitive and non-cognitive skills. To our knowledge no other economic study has tried to address this question so far.

In the remaining of the paper we first discuss the skills production function and the assumptions needed to identify the causal effect of computer use on cognitive and non-cognitive skills. We rely on a rich set of controls available in the LSAC, and assess the robustness of our results by comparing alternative estimators. Our results indicate that children using computers are more likely to score high in cognitive skill tests but less well in terms of non-cognitive skills. Computer use matters mainly during the weekend, and the effects are larger for girls and for children with highly educated parents.

The paper unfolds as follows: In section 4.2 we review the main findings of the computer literature. Section 4.3 introduces the skill production function and then discusses the identification of the parameter of interest. Section 4.4 presents the cohort data that we use. Section 4.5 presents our findings. Section 4.6 concludes.

4.2 Literature

The literature evaluating the impact of computer access and use on children's outcomes is still quite limited, probably due to the fact that computers entered schools and houses on a large scale only in the last 10 to 15 years. In this section we give a short summary of those studies evaluating the effect of computer use on labor market outcomes, educational attainment and cognitive skills.

4.2.1 Effect of Computer use on Labor Market Outcomes

Krueger (1993) uses Current Population Survey data to examine whether workers who use a computer at work earn a higher wage rate than otherwise similar workers who do not use a computer at work. Given the cross-section data, he estimates the causal effect through OLS where identification relies on a rich set of controls, including 2 digits occupational sectors. Estimates suggest that workers who use computers on their job earn 10 to 15 percent higher wages.

DiNardo and Pischke (1997) revisit Krueger's analysis and investigate whether his estimates reflect a true return to computer skills or just selection: i.e. higher wage workers use computers on their jobs. They do so using three large cross-sectional surveys from Germany. Like in Krueger's paper, they estimate an OLS model where identification relies on a rich set of controls. They find that the estimated wage differential associated with computer use in Germany is very similar to the U. S. differential. However, they also find large differentials for on-the-job use of calculators, telephones, pens or pencils, or for those who work while sitting down. They conclude that these returns to office tools, including computers, are probably driven by substantial selection.

4.2.2 Effect of Computer use on Educational Attainment

Schmitt and Wadsworth (2006) explore the link between ownership of a home computer at ages 15 and 17 and subsequent educational attainment in the principal British school examinations taken at ages 16 (GCSEs) and 18 (A-levels). Using the British Household Panel Survey (BHPS), they estimate the causal effect using a probit model where identification relies on a rich set of controls such as household income, mother and father's education, mother and father's age and number of dependent children living in the household. The data show a significant positive association between PC ownership and the qualifications obtained. The frequency of PC use also appears to be weakly correlated with positive educational outcomes at age 16.

Beltran, Das, and Fairlie (2006) look into the relationship between computer ownership and high school graduation in the US, using recent panel data from matched CPS files and the NLSY97. Using a probit model with a rich set of controls they find that home computers are associated with a 6-8 percentage point higher probability of graduating from high school. They also estimate a bivariate probit model for the joint probability of computer ownership and high school graduation using parental use of the Internet at work and the presence of another teenager in the household as instruments. The bivariate probit leads to coefficient estimates that are similar to the original probit estimates, although statistically insignificant.

4.2.3 Effect of Computer use on Skills

Angrist and Lavy (2002) assess the short-run consequences of increased computer-aided instruction (CAI) technology in Israeli schools. The causal effect is estimated using an OLS model and a 2SLS where the IV is given by an Israeli Government program that funded a large-scale computerization effort in many elementary and middle schools. The schools that received support were more likely to use CAI. Their results do not support the view that CAI improves learning, at least as measured by pupil test scores. They find a consistently negative and marginally significant relationship between the programme induced use of computers and 4th grade Maths scores. For other grades and subjects, the estimates are not significant, though also mostly negative.

Rouse, Krueger, and Markman (2004) present results from a randomized study of a well-defined program of computers use in US schools (grade 3 to 6): a popular instructional computer program, known as Fast ForWord, which is designed to improve language and reading skills. They assess the impact of the program using four different measures of language and reading ability. The causal effect is estimated using an OLS model where identification relies on randomization: i.e. in selected schools some students were randomly assigned Fast ForWord. Their estimates suggest that while use of the computer program may improve some aspects of students' language skills, it does not appear that these gains translate into a broader measure of language acquisition or into actual reading skills.

Banerjee, Cole, Duflo, and Linden (2007) look at the results of a randomized experiment conducted in schools in urban India (grade 3 and 4). A computer-assisted learning program was randomly assigned to some schools for up to two years. They find that the program was very effective, increasing math scores by 0.36 standard deviations the first year, and by 0.54 standard deviation the second year. However, they find that the effect of the program decays fast after the program ends, but this result is common to another treatment that provided teacher support rather than computer-assisted learning.

Subrahmanyam et al. (2001) survey the psychology literature. Several studies provide preliminary evidence that computer use is positively correlated with academic achievement. Few studies have examined the effect of children's time on computers on their social skills and friendships. The existing research suggests that frequent game players actually meet friends outside school more often than less frequent players and no differences have been found in the social interactions of computer game players vs. non-players. However most of these results apply mainly to teenagers.

4.2.4 Effect of TV use on Skills

To conclude our literature review we summarize the main findings on the effect of TV time on children's skills.

Gentzkow and Shapiro (2008) look at the effect of preschool television exposure on standardized test scores later in life. Using heterogeneity in the timing of television's introduction as a source of identification, they find that an additional year of preschool television exposure raises average test scores by about .02 standard deviations. These effects are largest for children from households where English is not the primary language, for children whose mothers have less than a high school education, and for non-white children.

Schmidt and Anderson (2007) provide an overview of the findings in the psychology literature. Exposure to educational programs, such as Sesame Street, has positive effect on children's vocabulary learning and this effect is long lasting. They do not find evidence that TV displaces intellectually valuable activities. In fact TV replaces activities similar to TV viewing such as radio listening, comic book reading and moviegoing.

4.3 The Production Function

In our data we observe the children at two points in time, when they are aged 4/5 (2004) and 6/7 (2006). Since it is unlikely that they made extensive use of a computer before age 4, let us start with a simple two period model $t = 1, 2$. Denote by C_t computer time at time t, by FI_t a vector of family inputs, by SI_t a vector of school inputs and by OM_t time spent using other media devices such as TV and video games. Let also μ denote children's unobserved time constant endowments (like innate abilities). Here μ is not 1 dimensional but rather a vector including a range of cognitive and non-cognitive innate abilities. Finally denote by T_{jt} the j^{th} test score measured at time t and by ϵ_t the measurement error in T_{jt}. As well as for μ, there is a vector of test scores T that can summarize the main cognitive and non-cognitive skills.

4.3.1 Period 1

The production function of each test score in period 1 can be written as:

$$T_{j1} = g_j(C_1, FI_1, SI_1, OM_1, \mu, \epsilon_1) \tag{4.1}$$

where we are assuming that any non-media input enters either FI_1 or SI_1. In this paper our parameter of interest is the effect of C_1 on T_{j1}, holding all other inputs constant. It is easy to see why the identification of this parameter is complicated by endogeneity problems. C_1 depends on the parental decision to own and make available a computer but also on the child decision to spend some time using it. Unobserved family, school and media inputs together with the child's innate abilities might be correlated with C_1 but also T_{j1}. Measurement error in C_1 can instead cause attenuation bias. In the data the parents were asked to report the time spent by their children using the computer. It is possible that some parents could only provide a rough guess. Therefore ϵ_1 can include measurement error in C_1. [2]

Todd and Wolpin (2003) discuss alternative estimation strategies under the assumption that the g function is linear, an assumption that we also make. Let X_1 denote observed family, school and other media inputs and let U_1 denote the unobserved ones.

$$T_{j1} = \alpha_{j1} + \beta_{j1}C_1 + X_1\gamma_{j1} + v_{j1} \tag{4.2}$$

where $v_{j1} = U_1\delta_{j1} + \mu\rho_{j1} + \epsilon_1$ ($\gamma_{j1}, \delta_{j1}, \rho_{j1}$ are vectors). The simplest way to estimate equation (4.2) is to use the OLS estimator and assume that we can control for the most important inputs influencing both C_1 and T_{j1} such that $E(v'_{j1}C_1) = 0$. The LSAC survey designers put a lot of care in collecting very detailed information regarding parental background, home and school care. In the results section we discuss what variables we can use to approximate the family, school and other media inputs. Yet even rich data can rarely allow to control for the innate abilities of the child μ. One possibility is to assume that the parental decision to own a computer is not a function of μ. That is parents own a computer mainly for their work, internet browsing or other personal uses so that the ownership decision does not depend on the children's characteristics. [3] If this is the case, and there are no other unobservable entering both the parental decision and the production function, than computer ownership HC_1 can serve as an instrument for C_1 since $E(v'_{j1}HC_1) = 0$ but $E(HC'_1C_1) \neq 0$. [4] Using an IV is also the only way to solve the measurement

[2]Test scores are the best available proxy of true skills, but they are still likely to measure these skills with errors. Thus skills's measurement error might also enter ϵ_1.

[3]In the data, parents were not asked whether they had a home computer but rather whether the child had access to one. Therefore parents had to take two decisions: whether to own a home computer and whether to make it available to the child. The latter could be correlated with μ. From the data, we only know whether the child had access to a computer at home. However, since in wave 1 (wave 2) 77% (88%) of the children had access to one, it is unlikely that many parents had a computer but did not make it available. That reduces the choice space to a simple ownership decision.

[4]One argument against HC_1 satisfying the exclusion restriction is time displacement. Since the parents own

error problem. Angrist and Imbens (1995) show that under heterogenous treatment effects, the IV estimator will identify the LATE, that is the return for those children who actually use the computer (since some children have access to a computer but do not use it, so that C_1 is not affected by the instrument). In period 1 it is also possible to test the robustness of OLS estimates by including a future measure of computer use C_2 in equation (4.2). Conditional on C_1, future computer use should not be correlated with T_{j1} unless μ or U_1 are correlated with C_2. [5]

4.3.2 Period 2

The production function in period 2 is:

$$T_{j2} = g_j(C_{2:1}, FI_{2:1}, SI_{2:1}, OM_{2:1}, \mu, \epsilon_2) \tag{4.3}$$

where the subscript $2:1$ indicates that we include both period 2 and 1 inputs. Every input of the production function at time 2 can have an effect on T_{j2} through its contemporaneous or lagged level, that is we do not restrict past inputs to drive T_{j2} only indirectly through the current inputs. This is true also for computers where use in period 1 (age 4/5 in our data) might have permanent effects on the test scores besides the effect on C_2. In other words we are interested in the timing of these inputs. If we only include C_2 its coefficient would pick up the effect of the whole computer history but we would not know when this input is most effective. According to Cunha et al. (2006) the timing of inputs matters because some skills can be shaped only when children are very young. Once again we assume that the production function is linear in its inputs:

$$T_{j2} = \alpha_2 + C_{2:1}\beta_{j2} + X_{2:1}\gamma_{j2} + v_{j2} \tag{4.4}$$

where $v_{j2} = U_{2:1}\delta_{j2} + \mu\rho_{j2} + \epsilon_2$. Therefore in equation (4.4) we are interested in estimating β_2 which is a 2×1 vector. The estimation of this equation is once again plagued by endogeneity

a computer, presumably they spend some time using it. If parental computer time displaces time with the child, or time otherwise invested in producing T_{j1} inputs that we do not control for, then $E(v'_{j1}HC_1) \neq 0$. However, if parental computer time displaces "unproductive" time, for instance parental TV time, then the exclusion restriction holds.

[5] However, even if $E(C'_2\mu) = E(C'_2U_1) = 0$, C_2 might be correlated with ϵ_1. This would happen if C_2 is a function of previous period test scores T_1.

problems potentially even more severe since now we are interested in the causal effect of the two endogenous variables C_1 and C_2. Besides OLS, Instrumental Variable estimation is still possible using HC_1 and HC_2 as instruments provided they are not multicollinear. However, consistency of the IV estimator now requires very strong restrictions on the time 2 parental decision. For $E(v'_{j2}HC_2) = 0$ to hold, the parental decision to own a computer must be uncorrelated with C_1 and T_1, since these are a function of μ, and T_1 is also a function of U_1. Todd and Wolpin (2003) discuss the estimation of the production function (4.4) using the Value Added model. The idea is to include a lagged test score T_{j1} on the right hand side. Intuitively, since the lagged test score is a function of μ, including it among the control variables one might reduce the omitted variable bias. However Todd and Wolpin (2003) also show that the Value Added model solves the endogeneity problem only if the impact of the ability endowment μ declines over time at a rate equal to the first order correlation across test scores. [6] Finally it is also possible to estimate the production function through the First Difference (or Fixed Effect) estimator. This estimator relies on other strong assumptions. The first two terms of v must be time constant, that is $(U_{2:1}\delta_{j2} + \mu\rho_{j2}) - (U_1\delta_{j1} + \mu\rho_{j1}) = 0$. Even if $C_{2:1}$ was orthogonal to $U_{2:1}$, the ability endowment must have a constant effect over time, $\rho_{j2} = \rho_{j1}$. In principle there is no reason why this should be the case and this equality holds for all the cognitive and non-cognitive abilities in the μ vector. Also, First Difference requires strict exogeneity. However this would be violated whenever C_2 is a function of T_1 either through the parental or children choice functions, since in that case $E(C'_2\epsilon_1) \neq 0$.

Later in the paper we provide estimates of the linear production functions in period 1 and 2. There are two main reasons why we estimate both functions rather than just the one in period 2. First, we are interested in the determinants of cognitive and non-cognitive skills because they, in turn, will act as determinants of educational choices and labor market outcomes. As much as both C_1 and C_2 might enter the period 2 production function, with C_1 still having a direct effect

[6] This can be easily seen under linearity. Using equations (4.2) and (4.4), and letting $'$ and $''$ indicate the first and second element of the vectors β, γ, δ:

$$\begin{aligned} T_{j2} - \phi T_{j1} =& (\alpha_2 - \phi\alpha_1) + \beta''_2 C_2 + (\beta'_2 - \phi\beta_1)C_1 + X_2\gamma''_2 + X_1(\gamma'_2 - \phi\gamma_1) \\ &+ U_2\delta''_2 + U_1(\delta'_2 - \phi\delta_1) + (\rho_2 - \phi\rho_1)\mu + \epsilon_2 - \phi\epsilon_1 \end{aligned}$$

This also shows that U_2 and U_1 still enter the error term unless $\delta''_2 = (\delta'_2 - \phi\delta_1) = 0$. Moreover, C_2 will be correlated with ϵ_1 if previous test scores enter the parental or children choice functions.

conditional on C_2, then we can also imagine a schooling or wage function where both T_1 and T_2 enter as inputs. If some learning processes, investments or choices are made at very young ages, T_1 might have a role even conditional on T_2. For this reason we are interested in the production functions of both T_1 and T_2. Second, in the data the vector of cognitive skill scores between period 1 and 2 is not the same, since some tests are age specific. Therefore the outputs of the production functions are not identical in the two periods. We refer to the data section for a more complete explanation of the cognitive skills measures.

4.4 Data

The data comes from the Growing Up in Australia, the Longitudinal Study of Australian Children (LSAC). This study aims to examine the impact of Australias unique social and cultural environment on the next generation. During 2004, over 10,000 children and their families were recruited to the study from a sample selected from the Health Insurance Commissions Medicare database. It is intended that these children and their families will be interviewed biannually until 2010, and possibly beyond. During 2004, the study recruited a sample of 5,107 infants (children born March 2003-February 2004) and 4,983 children aged 4-5 years (children born March 1999-February 2000) in a dual cohort cross-sequential design. Data for the first two waves of each cohort are now available. In what follows we focus on the older cohort, aged between 4 $\frac{1}{2}$ and 6 $\frac{1}{2}$ at the time of the two surveys. We then create our sample by selecting those children for whom data were collected at both waves.

4.4.1 Computer Access and Use

There are a number of variables that measure computer access/use by the child. In Wave 1 parents were asked whether the study child had access to a computer at home and if so, how many hours the child used the computer on a typical weekday and on a typical weekend day. Unfortunately in Wave 1 the number of hours were recorded in bands and not in continuous form. [7] Parents were also asked about the number of televisions at home and how many hours

[7] The 5 bands were coded as follow: 1. Five or more hours; 2. From three to five hours; 3. From one to three hours; 4. Less than one hour; 5. No use.

the child spent watching TV (still distinguishing between a weekday and weekend, and with hours coded in bands). If the child attended school, interviewers would interview the school's teacher, subject to parental authorization. Teachers were then asked whether the school was equipped with computers and how often the children used them. [8]

In Wave 2 parents were asked the same questions though this time computer and TV use were recorded as continuous variables. Moreover, in this second wave parents were also asked whether the children had access to a video game console such as Xbox, Playstation or Nintendo and if so, how many hours (weekday/weekend) they spent using it.

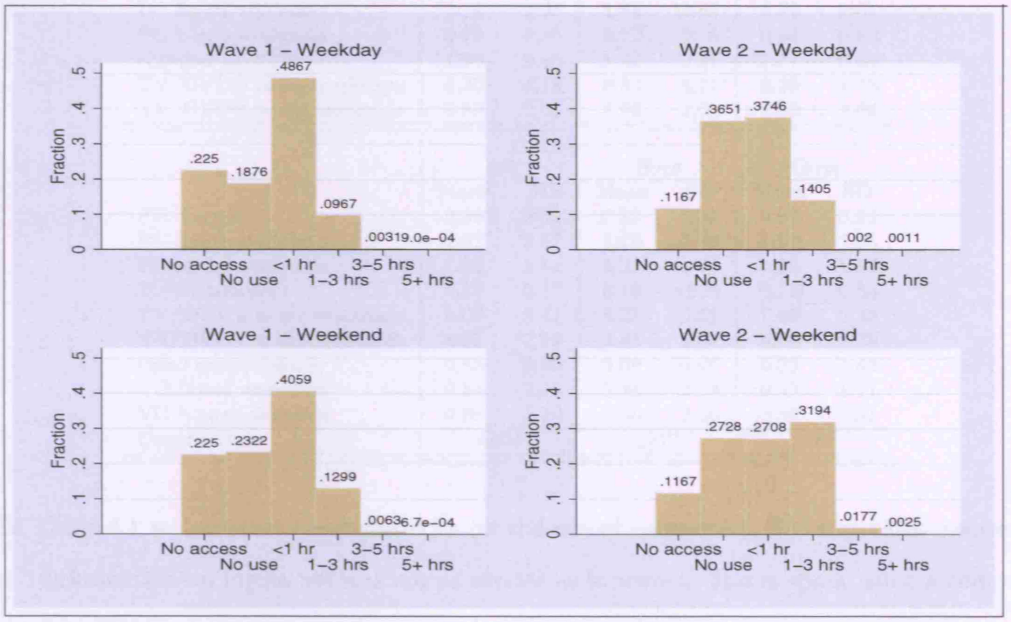

Figure 4.2: Home Computer Access and Use

Figure 4.2 shows computer access and use in waves 1 and 2. We distinguish between children who had no access to a computer (No access), those who had access but did not use it (No use) and those who had access and spent some time using it, where computer time is coded in 4 discrete hour bands (<1hr, 1-3hrs, 3-5hrs, 5+hrs). Children were more likely to have access to a computer in wave 2. By then only 11.67% of children could not access one. Perhaps parents decided to let the children use the computer as the children became older and started school

[8]Provided the school was equipped with computers, frequency of use was coded as follows: 1. Two or more hours per day; 2. From one to two hours per day; 3. Less than one hour per day; 4. A few times a week; 5. A few times a month; 6. Less often; 7. Never.

or it could be simply the result of the upward trend in computer ownership that we have seen in figure 4.1. [9] On the other hand quite a large fraction of 6/7 years old children did not use a computer even if they had access to one. If we look at weekdays, the figure suggests that as they aged, children either did not use the computer at all or became heavier users of it. There is instead a clear increase in computer use during the weekend between the two waves.

Table 4.1: Media Access and Use

	Wave 1					
	All		Boys		Girls	
	Mean	SD	Mean	SD	Mean	SD
PC Access	0.77	0.41	0.77	0.42	0.77	0.41
PC hours weekdays	1.78	2.28	1.84	2.51	1.72	2.01
PC hours weekends	0.70	0.85	0.72	0.89	0.68	0.80
Number of TV's	1.80	0.85	1.82	0.87	1.77	0.84
TV/DVD's hours weekdays	8.30	4.73	8.33	4.71	8.26	4.75
TV/DVD's hours weekends	3.89	2.06	3.90	2.09	3.88	2.02
	Wave 2					
	All		Boys		Girls	
	Mean	SD	Mean	SD	Mean	SD
PC Access	0.88	0.32	0.88	0.31	0.88	0.32
PC hours weekdays	1.67	2.67	1.76	2.68	1.57	2.65
PC hours weekends	1.21	1.51	1.31	1.65	1.10	1.33
TV in bedroom	0.17	0.37	0.19	0.39	0.14	0.35
TV/DVD's hours weekdays	8.07	5.41	8.25	5.51	7.89	5.30
TV/DVD's hours weekends	4.81	2.89	4.93	2.97	4.69	2.79
video game console	0.53	0.49	0.68	0.46	0.39	0.48
VG hours weekdays	0.84	2.19	1.34	2.78	0.33	1.11
VG hours weekends	0.98	1.70	1.55	2.06	0.39	0.91
Observations	4464		2277		2187	

In Table 4.1 we presents basic statistics on the use of computers, TV and video games. In wave 1 information on media use was not as precise as in wave 2. Hours spent using a computer or watching TV were coded in bands, parents were not asked about video games and we know the total number of TV's in the house but not whether the child had one in his/her own bedroom. In order to construct the figures in table 4.1 (Wave 1) we recoded number of hours in continuous form. For both computer and TV hours, we used the median number of hours within each band from wave 2 and imputed that figure for wave 1 observations. [10] The average child was using the computer for a *total* of 1.78 hours during the week, and a *total* of 0.70 hours during the weekend. Boys spent more time than girls using it, but this difference is not very large. Importantly, there is evidence of variation over time. Not reported, the correlation between C_1 and C_2 is equal to

[9] It is not possible to disentangle this two effects with data from only one cohort. Nevertheless as soon as the younger LSAC cohort reaches age 4/5, we should be able to say more.

[10] Say that in wave 2 the median number of minutes for those children in the '1 to 3 hours' was 150, then we would impute 150 minutes also for those children that in wave 1 fall within this '1 to 3 hours' band.

0.25. If C_1 and C_2 were to be multicollinear, estimation of equation (4.4) would be problematic resulting in large standard errors. In the appendix (table 4.12) we illustrate computer access/use variation over time, using the same coding as in figure 4.2. Children also spent 8 hours watching TV during the weekdays and almost 4 hours during the weekend. Once again boys stayed slightly longer than girls in front of a TV. In wave 2, when children were aged between 6 and 7 years old, almost everyone had access to a computer at home. Compared to wave 1, children used it less during the weekdays but more during the weekend, and a similar pattern exists also for TV use. Since in wave 2 children were aged 6 to 7 years old, and therefore all enrolled in school, it is possible that they had less home time during the week. Parents were also asked whether the child had his/her own TV in the bedroom. Almost one in five children had one. However no information was collected about the number of televisions at home. Finally in wave 2 more than half of the children had a video game console but boys spent remarkably more time than girls using it. Overall, at wave 2 an average (median) boy spent around 19 (17) hours using a combination of computer, TV and video games, while an average (median) girl spent 16 (14) hours.

Not reported in the table, the LSAC data also show that in wave 2 (wave 1) 81% (70%) of the schools had a computer in the classroom, though sample sizes are smaller, since not all children went to school (mainly in wave 1) or because the parents did not authorize the interviewer to go to the school.

Unfortunately the LSAC data does not contain information on what the children used the computer for. However, in 2006 the Australian Bureau of Statistics (ABS) has conducted a survey of 'Children's Participation in Cultural and Leisure Activities', which includes information details on children's use of computers and the internet for different age groups. In table 4.2 we report the main statistics. Children in the five to eight years group, our LSAC reference group, used the computer mainly to play games, followed by school or educational activities. The table also illustrates that as children age, less time is devoted to playing games while more and more time is spent in internet activities like browsing or emailing.

Table 4.2: Home Computer Usage, Activities

	5-8 yrs	9-11 yrs	12-14 yrs
Emailing or messaging	28.6	53.5	69.1
Other internet based activities	7.9	30.4	57.3
Playing games	87.7	80.7	69.9
School or educational	62.0	83.7	92.5
Other activities	3.3	2.9	3.8

Source: ABS. Study 4901.0 - Children's Participation in Cultural and Leisure Activities, Australia, Apr 2006.
Numbers in the table give the proportion of children carrying on that activity.

4.4.2 Cognitive and Non-Cognitive Skills

The LSAC children were administered three cognitive skill tests depending on their age.

- **Who am I? Test** (Wave 1 only) The Who am I? is a direct child assessment measure that requires children to copy shapes (circle, triangle, cross, square, and diamond) and write numbers, letters, words and sentences. One item was added to the standard Who Am I? booklet for use in LSAC. It is used for the children at ages 4 to 5 years to assess the general cognitive abilities needed for beginning school.

- **Peabody Picture Vocabulary Test** (Waves 1 and 2) A short form of the Peabody Picture Vocabulary Test (PPVT - III), a test designed to measure a child's knowledge of the meaning of spoken words and his or her receptive vocabulary for Standard American English. This adaptation is based on work done in the United States for the Head Start Impact Study, with a number of changes for use in Australia. The Wave 1 and Wave 2 versions of the PPVT contain different, although overlapping, sets of items of appropriate difficulty for children aged 4-5 years and 6-7 years. A PPVT stimulus book with 40 plates of display pictures was used. The child is not required to define words but to show what they mean by pointing to (or saying the number of) a picture that best represents the meaning of the word.

- **Matrix Reasoning Test** (Wave 2 only) Children completed the Matrix Reasoning (MR) test from the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV). This test of non-verbal intelligence presents the child with an incomplete set of pictures and requires them to select the picture that completes the set from 5 different options.

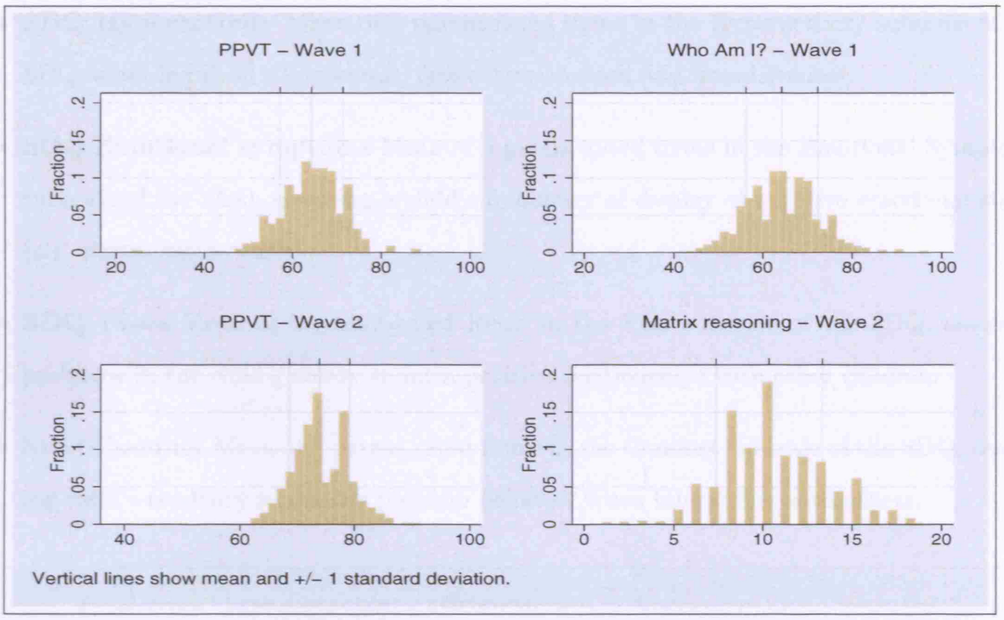

Figure 4.3: Cognitive Skills

Figure 4.3 shows the distribution of the cognitive test scores. Each distribution is quite symmetric. The matrix reasoning score has a different scale from the other tests. Later we standardize each test score to have mean zero and standard deviation 1.

In the LSAC, non-Cognitive skills are measured through both parental and teacher assessment. In the two waves parents and teachers were asked 25 questions about children's behavior. However, teachers' answers are available only if the child went to school and provided the parents authorized the interviewer to go to the school. Because of the larger sample size and in order to avoid sample selection (in school) problems, in the remaining of the paper we only use parental assessment. Most of the 25 questions did not change between the two waves and are described in the Appendix. From their answers LSAC data managers constructed five indicators of these skills.

- **SDQ Prosocial** Mean of 5 parent-rated items in the Prosocial subscale of the Strengths and Difficulties Questionnaire (SDQ), assessing the child's propensity to behave in a way that is considerate and helpful to others.

- **SDQ Hyperactivity** Mean of 5 parent-rated items in the Hyperactivity subscale of the SDQ, assessing child's fidgetiness, concentration span and impulsiveness.

- **SDQ Emotional symptoms** Mean of 5 parent-rated items in the Emotional Symptoms subscale of the SDQ, assessing a child's frequency of display of negative emotional states (e.g. nervousness, worry).

- **SDQ Peers** Mean of 5 parent-rated items in the Peer subscale of the SDQ, assessing problems in the child's ability to form positive relationships with other children.

- **SDQ Conduct** Mean of 5 parent-rated items in the Conduct subscale of the SDQ, assessing child's tendency to display problem behavior when interacting with others.

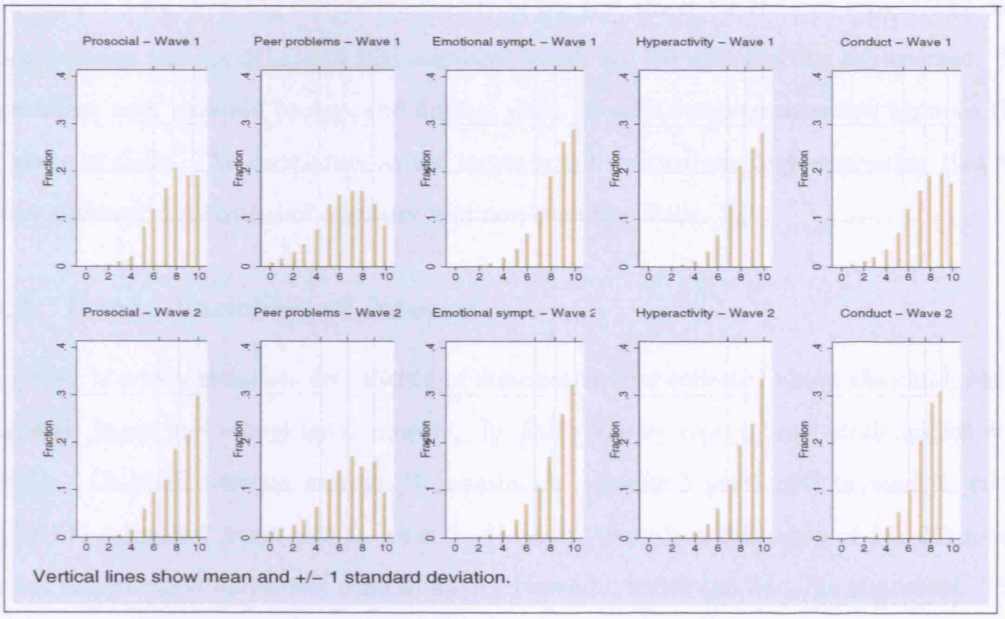

Figure 4.4: Non-Cognitive Skills

The non-cognitive scores are ordered such that a higher score corresponds to less behavioral problems, i.e. better non-cognitive skills. Figure 4.4 shows the distribution of the non-cognitive scores. These scores have right-skewed distributions (that is the majority of children do not have behavioral problems) and less variation than the cognitive scores. Non-cognitive scores are also standardized to have mean zero and standard deviation 1 before the estimation.

Table 4.3: Skills Correlation Matrix

	ppvt	wai	soc	hypr	emot	peer	cond
			Wave 1				
ppvt	1.00						
wai	0.29	1.00					
soc	0.08	0.13	1.00				
hypr	0.20	0.24	0.34	1.00			
emot	0.13	0.04	0.11	0.20	1.00		
peer	0.16	0.11	0.26	0.24	0.38	1.00	
cond	0.12	0.11	0.38	0.46	0.28	0.26	1.00
			Wave 2				
ppvt	1.00						
matrx	0.28	1.00					
soc	0.06	0.03	1.00				
hypr	0.11	0.15	0.32	1.00			
emot	0.07	0.04	0.10	0.23	1.00		
peer	0.07	0.04	0.25	0.29	0.41	1.00	
cond	0.10	0.09	0.38	0.48	0.33	0.33	1.00

In table 4.3 we show the correlation matrix of the test scores for the two waves. The correlations between the cognitive and non-cognitive scores are low and positive as expected. This is consistent with parental background driving skills, or with complementarities between these two kinds of skills. The correlation across scores is not particularly high suggesting that they capture different dimensions of cognitive and non-cognitive skills. [11]

4.4.3 Other Variables of Interest

The LSAC is a very rich data set. Plenty of information was collected about the child, his/her household, home and school environments. In Table 4.4 we report basic statistics for a few variables. Children were on average 57 months old (almost 5 years old) in wave 1, and 83 months old (almost 7 years old) in wave 2. However there is a difference of 18 (22) months between the youngest and oldest child in wave 1 (wave 2), which can be quite important. These children had on average 1 $\frac{1}{2}$ siblings and in 95% of the case there were at most 3 siblings. The average mother was slightly younger than 30 years old at birth, and the average father slightly older than that. Most parents had some educational qualification beyond year 12 (high school). Father's income was substantially larger than mother's income, also due to a low fraction of mothers working full-time. The fraction of mothers working either full or part time rose between

[11] To investigate this point further we also run a factor analysis using the principal-components factor method. At both waves we find three main factors. Upon rotation, the factor loadings suggest the following grouping: (1) PPVT and WAI scores; (2) SDQ Prosocial, Hyperactivity and Conduct scores; (3) SDQ Emotional symptoms and Peers scores.

wave 1 and 2.

Table 4.4: Other Variables of Interest

	Wave 1		Wave 2	
	Mean	SD	Mean	SD
Child's Age (months)	57.40	2.62	83.66	2.97
Number of Siblings	1.47	1.02	1.58	1.03
Father Age (years)	37.50	5.86	39.48	5.97
Mother Age (years)	34.92	5.43	37.02	5.45
Father Higher Education	0.75	0.43	0.76	0.42
Mother Higher Education	0.64	0.47	0.68	0.46
Father Income (10 thous)	5.35	3.81	6.37	4.61
Mother Income (10 thous)	2.26	2.02	2.79	2.62
Mother Empl. Full-Time	0.20	0.40	0.25	0.43
Mother Empl. Part-Time	0.37	0.48	0.40	0.49

4.5 Results

In this section we provide estimates of the linear production functions in equations (4.2) and (4.4). Given the endogeneity problems discussed in section 4.3, and given that all estimators demand relatively strong assumptions, in what follows we report the parameters of interest of (4.2) and (4.4) using different estimators. All the test scores have been standardized to have mean zero and standard deviation 1. Computer time is measured as *total* weekly hours.

4.5.1 Period 1 Estimation

In table 4.5 we present the estimated effects of computer use on cognitive and non-cognitive skills. The first column (OLSa) illustrates the estimated impact when controlling for some measures of family, school and other media inputs such as weekly hours in child care, indoor and outdoor activities involving a family member, hours spent watching TV. [12] Children using the computer more often score higher in the Peabody Picture Vocabulary and Who am I? tests. With regard to the non-cognitive scores, the SDQ Prosocial, Hyperactivity and Conduct coefficients are positive and significant at 1% level. Their positive coefficients indicate that children using a computer have better non-cognitive skills.

In column OLSb we add a rich set of household characteristics such as parental education,

[12] A full list of the control variables used in this and later tables is available in the Appendix. Note that in table 4.5 media activities are given by the number of hours watching TV/DVD's while for wave 2 we also include hours playing with videogame consoles.

Table 4.5: Production Function - Period 1

	OLSa	OLSb	OLSc	IVa	R^2/N	Future	R^2/N
Peabody Pict. Vocabulary	0.024**	0.022**	0.022**	0.053**	0.21	-0.008	0.22
	(0.005)	(0.005)	(0.005)	(0.011)	3990	(0.011)	880
Who am I?	0.031**	0.029**	0.029**	0.045**	0.27	0.017	0.25
	(0.005)	(0.005)	(0.005)	(0.010)	4396	(0.009)	979
SDQ Prosocial	0.016**	0.015**	0.015**	0.030**	0.07	-0.023*	0.07
	(0.005)	(0.005)	(0.005)	(0.012)	4453	(0.011)	999
SDQ Hyperactivity	0.013**	0.010	0.010	0.019	0.13	-0.014	0.13
	(0.005)	(0.005)	(0.005)	(0.011)	4453	(0.010)	999
SDQ Emotional symptoms	0.006	-0.000	0.000	0.008	0.07	0.002	0.06
	(0.005)	(0.005)	(0.005)	(0.011)	4452	(0.011)	999
SDQ Peers	0.009	0.005	0.006	0.030**	0.10	0.004	0.07
	(0.005)	(0.005)	(0.005)	(0.011)	4453	(0.011)	999
SDQ Conduct	0.018**	0.013*	0.013*	0.038**	0.08	-0.021	0.08
	(0.005)	(0.005)	(0.005)	(0.011)	4453	(0.011)	999

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.
OLSa: control for type of school, child home and outdoor activities with family members, child extra activities such as sport and music classes, computer in school, home TV time.
OLSb: like OLSa plus control for household demographics, parental education and financial situation.
OLSc: like OLSb minus parental income and other indicators of financial distress.
IV: like OLSb plus instrument computer time with computer access.
Future: like OLSb but explanatory variable is future computer time C_2 rather than C_1 and include only those children who did not have access to a computer in period 1.

income, number of siblings etc. None of the additional controls is a direct family (FI), school (SI) or other media (OM) input but we rather consider them as important determinants of these inputs. Since it is quite rare to observe all inputs, household characteristics are often used as proxies in similar studies. Overall the coefficients are smaller but with no large change, though among the non-cognitive scores only the SDQ Prosocial coefficient is now significant at 1% level. This result is quite reassuring and it suggests that our set of inputs is quite comprehensive. Some researchers have criticized the use of household characteristics, and particularly of parental income, as a proxy of family or school inputs. They argue that an increase in the amount of an input holding income constant must imply a reduction in expenditures on other inputs. This could cause a misinterpretation of the coefficients. In column OLSc we then present the coefficients when excluding parental income and other indicators of financial distress from the set of control variables. The results are virtually identical to those in column OLSb. In the remaining of the paper we include parental income and other indicators of financial distress among the household characteristics.

To learn whether the effect is large or not we compare the computer coefficient in column

OLSb to those of TV/DVD and child care weekly hours (not reported in the table). [13] For the Peabody Picture Vocabulary, Who am I? and SDQ Prosocial scores, the TV/DVD coefficients are, in order of test score, -0.002, -0.008** and -0.008**, i.e. smaller and of opposite sign to the computer ones. The child care coefficients are -0.004, 0.008** and -0.001, again much smaller than the computer coefficients. Clearly endogeneity problems might bias these latter coefficients as much as the computer coefficient. However, unless the bias is large and possibly of different sign (i.e. computer coefficients are upward biased while TV/DVD and child care coefficients are downward biased) there is evidence that computer time is an important input in the production function.

Next we move to the IV estimator. Since we are not aware of any institutional change (laws or similar) that might affect C_t our approach is to use computer access at home (HC_1) to instrument C_1. In section 4.3.1 we discussed under what conditions this estimator is consistent. To satisfy the exclusion restriction we need computer access to be uncorrelated with unobserved inputs and the ability endowment. Given the large fraction of children with access to a home computer, we expect that, if anything, only a few parents owning a computer deny access to their children. Therefore it is unlikely that HC_1 is correlated with μ. In table 4.13 we then compare households with and without a computer over a number of observable characteristics. Households with a computer are on average older, better educated, richer and more likely to have the mother employed. Our assumption is that conditional on these and the other controls included in OLSb, households with and without a computer do not differ over any other unobserved input of the production function. With regard to the rank condition, a first stage regression of C_1 on HC_1 and all other control variables used in OLSb show that HC_1 coefficient is positive and significant (see table 4.14, in the appendix). By definition the HC_1 coefficient is simply equal to $E(C_1|X_1, HC_1 = 1)$. Back to table 4.5, we see that under the IV estimator the return to computer use becomes larger for all cognitive and non-cognitive scores. This result is consistent with attenuation bias caused by measurement error in C_1 but not with omitted variable bias caused by unobserved innate abilities, which, at least in the case of cognitive scores, is expected to drive the coefficient upwards. It is also possible that the measurement error bias more than compensate the omitted variable bias, or that the IV estimator identifies a LATE.

[13] Child care hours are given by the average weekly hours in school, kindergarten, pre-school or day care.

The fifth column (R^2/N) reports the adjusted R^2 for the richest OLS regression (OLSb) and the sample size (N). The R^2 is larger for the cognitive scores production function.

Finally, we run a robustness test by estimating the effect of C_2 on T_1 (Future column). As discussed in section 4.3.1, conditional on C_1, future computer use C_2 should have no correlation with T_1 unless C_2 is a function of μ (skill endowments) and U_1 (unobserved inputs). If that was the case it is likely that $E(v'_{j1}C_1) \neq 0$. In order to properly control for current computer use we select only those children who did not have access to a computer in the first period. [14] In our sample 14.83 % of the children gained access to a computer between the two waves. Only in the case of the SDQ Prosocial score C_2 has a statistically significant effect. However, the C_2 coefficient is negative. If this is just an omitted variable bias, then the true β_{j1} is actually larger and not smaller than the OLS estimates.

To recap, both the OLS and IV estimator indicate that computer use in period 1 (age 4/5) has a positive statistically significant effect on the cognitive scores and on the SDQ Prosocial and Conduct scores, with the OLS coefficients being relatively large compared to those of other inputs. The OLS estimator passes the robustness check where we test for the effect of C_2 on T_{j1} for all scores but the SDQ Prosocial. However in this case, the negative coefficient suggests that the omitted variables might actually downward bias the estimates.

4.5.2 Period 2 Estimation

In table 4.6 we show the parameter estimates for the period 2 production function (equation 4.4). For every test score function we report the C_2 (top) and C_1 (bottom) coefficients. The first three columns are obtained like in table 4.5 by controlling for family, school and other media inputs (OLSa), household characteristics (OLSb) and using computer ownership in both periods $HC_{2:1}$ to instrument $C_{2:1}$ (IV). The only difference is that for all OLS and IV estimators we now control for both periods characteristics ($X_{2:1}$) while in table 4.5 we controlled only for period 1 (X_1). Conditional on C_1, current computer use C_2 seems to have an effect only on the SDQ Peers score though the IV estimator is much smaller (in absolute value) and imprecise. The negative sign indicates that children spending more hours in front of the computer are more likely to have

[14] Alternatively, we could simply condition on C_1. However, C_1 might just be an imperfect proxy of current computer use.

Table 4.6: Production Function - Period 2

		OLSa	OLSb	IV	R^2/N	VA	R^2/N
Peabody Pict. Vocabulary	C_2	0.005	0.004	-0.003	0.21	0.005	0.34
		(0.004)	(0.004)	(0.017)	4409	(0.004)	3960
	C_1	0.021**	0.020**	0.028*	—	0.011*	—
		(0.005)	(0.005)	(0.014)	—	(0.005)	—
Matrix Reasoning	C_2	0.008	0.008	0.016	0.09	0.006	0.16
		(0.004)	(0.004)	(0.018)	4402	(0.004)	4347
	C_1	0.030**	0.027**	0.058**	—	0.021**	—
		(0.005)	(0.005)	(0.015)	—	(0.005)	—
SDQ Prosocial	C_2	-0.005	-0.007	0.003	0.07	-0.004	0.29
		(0.004)	(0.004)	(0.018)	4342	(0.004)	4333
	C_1	0.007	0.007	0.015	—	-0.001	—
		(0.005)	(0.005)	(0.015)	—	(0.005)	—
SDQ Hyperactivity	C_2	-0.000	-0.002	0.005	0.14	0.001	0.41
		(0.004)	(0.004)	(0.018)	4341	(0.004)	4332
	C_1	-0.002	-0.001	0.007	—	-0.006	—
		(0.005)	(0.005)	(0.015)	—	(0.004)	—
SDQ Emotional symptoms	C_2	0.001	-0.001	0.025	0.08	0.001	0.25
		(0.004)	(0.004)	(0.018)	4341	(0.004)	4331
	C_1	0.008	0.003	-0.018	—	0.003	—
		(0.005)	(0.005)	(0.015)	—	(0.005)	—
SDQ Peers	C_2	-0.012**	-0.014**	-0.003	0.10	-0.009*	0.25
		(0.004)	(0.004)	(0.018)	4341	(0.004)	4332
	C_1	-0.004	-0.007	-0.014	—	-0.009	—
		(0.005)	(0.005)	(0.015)	—	(0.005)	—
SDQ Conduct	C_2	-0.002	-0.006	0.023	0.10	-0.003	0.31
		(0.004)	(0.004)	(0.018)	4341	(0.004)	4332
	C_1	0.009	0.008	0.023	—	0.001	—
		(0.005)	(0.005)	(0.015)	—	(0.005)	—

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.
OLSa: control for type of school, child home and outdoor activities with family members, child extra activities such as sport and music classes, computer in school, home TV time.
OLSb: like OLSa plus control for household demographics, parental education and financial situation.
IV: like OLSb plus instrument computer time with computer access.
VA: like OLSb plus include lagged score on the right hand side.

peer problems. However C_1 has a positive effect on both cognitive scores, with the IV estimator being larger than the OLS one. The fifth column (R^2/N) report the adjusted R^2 for the richest OLSc regression and the sample size (N). The Peabody Picture Vocabulary function is the one with the largest R^2 while the Matrix Reasoning and other non-cognitive scores functions have a smaller R^2.

We then estimate the production function using the Value Added estimator (VA column). In section 4.3.2 we discussed the conditions under which this estimator is consistent. The estimates are obtained after augmenting the right hand side of each production function with the period 1 test score. Most of the C_2 coefficients drop, including the SDQ Peers coefficient, though the latter is still negative and statistically significant. The C_1 coefficients drop too, but this is expected since we are including the lagged score on the right hand side (see section 4.3.2, footnote 6). As we would expect the R^2 of the Value Added model is larger than the OLS estimator since the lagged score might be capturing the effect of unobserved innate abilities or past unobserved inputs. The sample size N is instead smaller since we only include those children for whom both periods scores are available.

We do not include the First Difference estimator instead mainly because the wave 1 computer hours were originally coded in bands. While it is already known that the First Difference estimator can exacerbate measurement error problems, in our case a ΔC variable created using our imputed continuous C_1 would generate even more measurement error. If instead we were to code both C_2 and C_1 in bands, we would loose all the children that did not change band between the two waves, roughly half of the sample.

To summarize, the results from period 2 suggest that computer use at young ages has a long lasting effect on cognitive skills, while current use has no strong effect. Per contra, C_2 has a negative effect on the SDQ Peers score, our indicator of the child's ability to form positive relationships with other children. That would be compatible with the hypothesis that children substitute time with other children with computer time. These results are consistent across the OLS, IV and Value Added estimators.

4.5.3 Weekday vs Weekend

We now try to exploit the information in our data by separating C_t into weekday (C_t^{wd}) and weekend (C_t^{we}) computer hours: $C_t = C_t^{wd} + C_t^{we}$. The C_t coefficient is expected to lie in between the C_t^{wd} and C_t^{we} ones. [15] In table 4.7 we show the results. We only present the coefficients for the three cognitive skills and the SDQ Prosocial, SDQ Peers and SDQ Conduct non-cognitive skills, and report the OLSb estimator, column OLSb(T_t), and the value added estimator, column VA(T_t).

Starting with the cognitive skills, we see that what is important is computer use during the weekend, with coefficients sensibly larger than those in tables 4.5 and 4.6. For the Matrix Reasoning test, C_2^{we} has now a statistically significant effect, even using the VA estimator, while in table 4.6 C_2 had a negligible effect. This is because the C_2 coefficient is a weighted sum of

[15] This is true if $Cov(C_t^{wd}, C_t^{ww}) > 0$, which is the case in our sample.

the C_2^{wd} and C_2^{we} ones. But why is it weekend computer time that matters? On the one hand since parents are more likely to be home (i.e. not working) during the weekend, they might be spending time with their children using educational software or other programs. On the other hand, it is possible that computer time during the weekday displaces other positive inputs of the cognitive production function, such as homework or other educational assignments, producing a zero sum effect, while during the weekend computer time displaces activities that are not cognitive skill enhancing, such that computer time has a net positive effect.

For the non-cognitive skills, whether it is the weekday or the weekend time that matters depends on the skill. In the case of the SDQ Prosocial function, it is weekday time that is important, though with an opposite sign between the two time periods. In the absence of information on computer activities and displaced activities, we do not have a clear intuition for this result. The SDQ Peers function shows instead a negative effect of C_2^{we}. If children using the computer during the weekend are less likely to form positive relationships with other children, it could be that computer time is displacing time playing and interacting with siblings or friends. It is also worth noting that the weekend coefficients have a larger standard error than the weekday ones.

4.5.4 Heterogeneity in the Production Function

In this section we investigate whether the production function parameters are heterogeneous. In particular we look at differences based on the children's sex and on their mothers's education and labor market status. Tables 4.8 and 4.9 illustrates the results for the two periods. For clarity we only report the OLSb estimator. Given the results in section 4.5.3, we also distinguish between C_t^{wd} and C_t^{we} and highlight the coefficients that were statistically significant in table 4.7.

For most scores, the impact of computer use is almost always larger for girls. There is some evidence that among teenagers, boys and girls use the computer differently, with boys spending more time playing games and girls using it more for emailing and chatting (see Subrahmanyam, Kraut, Greenfield, and Gross (2000)), though we do not know whether these differences in usage apply also to younger children. There is also evidence that boys and girls learn differently (see Gurian (2002)). However, a more complete investigation of these differences between boys and girls is beyond the scope of this paper.

Next, we divide our sample in three groups based on their mothers's education: below year 11, year 11 or 12 (completed high school), higher education. As we mentioned in the introduction, computer time might matter depending on the content and/or depending on the activities that are displaced by it. On the one hand, if it is content that matters, than children with better educated parents should have a higher return to computer time. This would be the case if better educated parents are more aware of which computer usages are educational or if they are more computer savvy themselves, and can teach their children how to use computers. On the other hand, if the effect comes mainly through the displaced activities, than children with low educated parents might have the highest return, since computer time might be more educational than time with parents. For instance, Bernal and Keane (2008) find that the effect of child care is positive mainly for children with low educated parents. The authors point at the displacement effect to explain this result. Tables 4.8 and 4.9 suggest that both channels might be in place. For most of the cognitive scores the effect of computer time is highest either for the low educated or the high educated parents groups.

Finally, we also test whether there is heterogeneity depending on the mother's working status. We distinguish between full-time, part-time and not working, the latter including mothers looking for a job, in maternity leave or out of the labor force. One way to explain a stronger return for weekend versus weekday time is to assume that parents can guide computer use better during the weekend, since they are more likely to be home and have time for the child. If this is true, than we would expect the difference between weekend and weekday use to be largest among children with working mothers. The last three columns of tables 4.8 and 4.9 indeed indicate that the difference between the weekend and weekday returns is larger for those children with mothers working full or part time. This result is also in line with the hypothesis that content matters.

4.5.5 Multimedia

Lastly, in this section we compare the return to computer, television and video games time. In Table 4.10 we present the result for the period 1 production function. The first columns correspond to the OLS return and is therefore identical to column OLSb in table 4.5. The second column (IV) differs from the one in table 4.5 because we now also instrument TV time with number of televisions in the house. That is we regress the scores on computer and TV time in the same regression and, for the IV estimates, instrument computer and TV time respectively with computer access and number of TV's in the house. The last two columns show the return to TV time.

Testing whether computer and TV time have a different return is interesting in light of our previous discussion. Both computers and TV are media devices, both will have an effect on children's skills depending on their content, on the activities they displace and on their intellectual stimulation. From table 4.10 it appears that computer and TV time have a very different effect. TV time has a statistically significant negative return an almost all scores, cognitive and non-cognitive. For both computer and TV time, the IV estimates are usually larger (in absolute value) than the OLS ones.

In table 4.11 we repeat the analysis for the period 2 production function. Now we also include, in the same regression, video game time, that is time spent playing games using consoles such as Xbox, Nintendo and Playstation. We only include period 2 media time and do not separate between current and lagged like we did in table 4.6 instead. This is because we do not observe video games time in period 1 but we want to keep the estimates comparable across the three media devices. [16] The IV estimates are obtained instrumenting computer, TV and video games time respectively with computer access, TV in child's bedroom (yes/no) and video game console access. It is unfortunate that in period 2 we do not observe number of televisions in the house. It is quite unlikely that the presence of a TV in the child's bedroom, as much as access to video game console, are uncorrelated with the child's characteristics. Nevertheless we still include the IV estimates for completeness. We also show the Value Added estimates obtained by including the lagged score on the right hand side. Television has still a negative effect for most of the scores, no matter which estimator is used, though the Value Added estimates are generally smaller. The effect of spending time playing with video games is also mostly negative even though only for the Peabody Picture Vocabulary score this effect is statistically significant.

[16] We do not include lagged computer, TV and video game time but we do include current and lagged measures for all other control variables.

4.6 *Conclusions*

The objective of this paper was to investigate the effect of using a home computer for children's cognitive and non-cognitive development. Data show that in OECD countries 70% or more of the households have a computer at home. Our Australian data also show that in families with young children this percentage can go up to almost 90% and that children do make use of computers even at very young ages. However not much is know about the effect of computers. Computers are a relatively new input in the children production function having entered the average household mainly in the last fifteen years. Previous research in economics has focused on the effect of computer on the adults' wage production function (with controversial findings), on high school graduation (positive effect) or on the effect of computer use in school, the latter often specific to a particular computer-assisted learning program (mixed findings). Psychologists instead have already completed some studies on the effect of home computer usage but data is mainly available for teenagers and some of these studies do not deal with the endogeneity of computer time.

In our work we use data from an Australian cohort born in 2000, with information collected in 2004 and 2006. The advantage of using this sample is twofold. These children are very young and data is recent. The latter is an important characteristic since computers, software, internet availability and parental computer's skills all have changed sensibly in the last two decades. We are not aware of any similar study.

For cognitive skills, our results indicate that computer time has a positive effect. The effect is long-lasting with early computer use showing an impact on test scores even two years later. This positive effect originates mainly from computer time during the weekend, is larger for girls, for children with low or highly educated parents and for children with working parents. The effect is large relatively to those of other inputs, such as child care, and is not shared by other media devices, such as television and video games which instead show a negative effect. This pattern of results suggests that the impact of computer time might be coming from different channels. First, by what computers are used for, i.e. content, since it is positive for children with highly educated parents and not for televisions and video games use. Second, because of the activities displaced by computer time, since the largest effect is found for children with low educated parents. Third, and again given the negative effect of television, because a computer is a very interactive device and therefore intellectually stimulating.

For the non-cognitive skills the evidence is more mixed, with the sign of the effect depending on the score and the age of the children. For the SDQ Prosocial score, which assess the child's propensity to behave in a way that is considerate and helpful to others, we find a positive effect for children aged between 4 and 5 years. This effect is larger for girls, and for children with highly educated and working parents. However two years later the effect turns negative. For the SDQ Peers score, which assesses the child's ability to form positive relationships with other children, we find a negative effect in period 2, mainly due to computer time during the weekend. It is harder to interpret the mixed effects on non-cognitive skills. A negative weekend effect is however consistent with the displacement of time spent in company of other children or adults. Nevertheless, a more exhaustive investigation of the mechanisms behind the computer effect would demand information on actual computer activities, which are not available in our data.

We test the robustness of our results using OLS, IV and Value Added estimators. Generally, the IV estimates are larger and the Value Added estimates lower than the OLS ones. However the pattern of result is quite consistent. We also test for omitted variable bias by testing whether future computer time has any effect on current score. Only in the case of the SDQ Prosocial score we find evidence of a correlation with unobserved characteristics, but if anything, the bias might be attenuating the effect rather then reinforcing it.

Table 4.7: Production Function - Weekday vs Weekend

Cognitive Skills					Non-Cognitive Skills				
		OLSb(T_1)	OLSb(T_2)	VA(T_2)			OLSb(T_1)	OLSb(T_2)	VA(T_2)
Peabody Pict. Vocabulary	C_2^{wd}	—	0.004	0.005	SDQ Prosocial	C_2^{wd}	—	-0.013*	-0.012*
		—	(0.006)	(0.005)			—	(0.006)	(0.006)
	C_2^{we}	—	0.007	0.006		C_2^{we}	—	0.010	0.017
		—	(0.011)	(0.010)			—	(0.011)	(0.010)
	C_1^{wd}	0.009	0.021**	0.012		C_1^{wd}	0.020*	0.009	-0.000
		(0.008)	(0.008)	(0.007)			(0.008)	(0.008)	(0.007)
	C_1^{we}	0.073**	0.014	0.006		C_1^{we}	0.001	-0.004	-0.009
		(0.022)	(0.022)	(0.021)			(0.023)	(0.024)	(0.021)
Who am I?	C_2^{wd}	—	—	—	SDQ Peers	C_2^{wd}	—	-0.006	-0.002
		—	—	—			—	(0.006)	(0.006)
	C_2^{we}	—	—	—		C_2^{we}	—	-0.031**	-0.024*
		—	—	—			—	(0.011)	(0.010)
	C_1^{wd}	0.015*	—	—		C_1^{wd}	0.003	-0.006	-0.007
		(0.007)	—	—			(0.008)	(0.008)	(0.008)
	C_1^{we}	0.078**	—	—		C_1^{we}	0.016	-0.006	-0.016
		(0.020)	—	—			(0.022)	(0.023)	(0.021)
Matrix Reasoning	C_2^{wd}	—	-0.005	-0.006	SDQ Conduct	C_2^{wd}	—	-0.011	-0.005
		—	(0.006)	(0.006)			—	(0.006)	(0.005)
	C_2^{we}	—	0.037**	0.034**		C_2^{we}	—	0.006	0.005
		—	(0.011)	(0.011)			—	(0.011)	(0.010)
	C_1^{wd}	—	0.016	0.014		C_1^{wd}	0.011	0.001	-0.004
		—	(0.008)	(0.008)			(0.008)	(0.008)	(0.007)
	C_1^{we}	—	0.063**	0.044		C_1^{we}	0.020	0.029	0.017
		—	(0.023)	(0.023)			(0.022)	(0.023)	(0.020)

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.

Table 4.8: Heterogeneity in the Production Function - Period 1

		Boys	Girls	M low Edu	M mid Edu	M high Edu	M work FT	M work PT	M not Wkg
Peabody Pict. Vocabulary									
	C_1^{wd}	0.013	-0.007	0.015	-0.001	0.017	-0.013	0.011	-0.002
		(0.010)	(0.014)	(0.017)	(0.019)	(0.011)	(0.016)	(0.011)	(0.016)
	C_1^{we}	0.049	0.109**	0.098	0.044	0.072**	0.120**	0.081*	0.059
		(0.029)	(0.036)	(0.072)	(0.054)	(0.028)	(0.042)	(0.035)	(0.042)
	N	2031	1959	542	856	2591	793	1502	1691
Who am I?									
	C_1^{wd}	0.008	0.021	-0.006	0.020	0.024*	0.008	0.014	0.028*
		(0.009)	(0.013)	(0.013)	(0.018)	(0.011)	(0.016)	(0.011)	(0.014)
	C_1^{we}	0.077**	0.094**	0.151**	0.028	0.066*	0.072	0.103**	0.036
		(0.026)	(0.032)	(0.056)	(0.051)	(0.026)	(0.044)	(0.032)	(0.036)
	N	2236	2160	599	938	2857	887	1650	1853
SDQ Prosocial									
	C_1^{wd}	0.018	0.031*	0.009	0.040	0.022	0.043*	0.030*	-0.007
		(0.011)	(0.014)	(0.016)	(0.022)	(0.012)	(0.017)	(0.013)	(0.016)
	C_1^{we}	-0.011	0.005	0.054	-0.019	-0.010	-0.013	-0.056	0.070
		(0.031)	(0.035)	(0.067)	(0.059)	(0.028)	(0.046)	(0.037)	(0.041)
	N	2268	2185	609	953	2889	898	1674	1875

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.

Table 4.9: Heterogeneity in the Production Function - Period 2

		Boys	Girls	M low Edu	M mid Edu	M high Edu	M work FT	M work PT	M not Wkg
Peabody Pict. Vocabulary									
	C_2^{wd}	0.003	0.013	-0.004	-0.001	0.007	0.003	0.006	0.007
		(0.009)	(0.009)	(0.017)	(0.020)	(0.007)	(0.016)	(0.008)	(0.012)
	C_2^{we}	0.008	-0.008	-0.060	0.009	0.005	0.011	-0.002	0.003
		(0.014)	(0.018)	(0.037)	(0.028)	(0.013)	(0.025)	(0.017)	(0.020)
	C_1^{wd}	0.021*	0.027	0.030	0.031	0.023*	0.018	0.010	0.052*
		(0.010)	(0.014)	(0.017)	(0.023)	(0.011)	(0.014)	(0.012)	(0.022)
	C_1^{we}	0.002	0.036	0.162*	-0.052	0.003	0.006	0.052	-0.038
		(0.029)	(0.036)	(0.079)	(0.061)	(0.027)	(0.045)	(0.035)	(0.046)
	N	2248	2161	551	843	3013	1122	1805	1481
Matrix Reasoning									
	C_2^{wd}	-0.004	-0.008	-0.021	-0.008	0.004	-0.020	-0.012	0.021
		(0.009)	(0.009)	(0.016)	(0.022)	(0.008)	(0.017)	(0.009)	(0.012)
	C_2^{we}	0.047**	0.026	-0.001	0.043	0.032*	0.059*	0.046*	-0.004
		(0.015)	(0.019)	(0.035)	(0.031)	(0.014)	(0.026)	(0.019)	(0.021)
	C_1^{wd}	0.010	0.030*	-0.004	0.033	0.031**	0.006	0.007	0.045*
		(0.010)	(0.015)	(0.016)	(0.025)	(0.012)	(0.015)	(0.013)	(0.023)
	C_1^{we}	0.039	0.109**	0.190*	-0.069	0.054	0.091	0.107**	0.025
		(0.032)	(0.038)	(0.076)	(0.067)	(0.030)	(0.048)	(0.038)	(0.048)
	N	2244	2158	552	843	3005	1117	1804	1480
SDQ Peers									
	C_2^{wd}	-0.001	-0.012	-0.025	-0.072**	0.005	-0.012	-0.000	-0.012
		(0.010)	(0.009)	(0.019)	(0.022)	(0.008)	(0.018)	(0.009)	(0.013)
	C_2^{we}	-0.039*	-0.028	0.024	0.010	-0.042**	-0.012	-0.028	-0.052*
		(0.016)	(0.018)	(0.041)	(0.031)	(0.014)	(0.027)	(0.018)	(0.022)
	C_1^{wd}	-0.012	0.010	-0.022	-0.014	0.002	0.018	-0.014	-0.014
		(0.011)	(0.014)	(0.019)	(0.026)	(0.012)	(0.015)	(0.012)	(0.024)
	C_1^{we}	-0.010	-0.016	0.169	0.047	-0.045	-0.065	0.030	0.008
		(0.033)	(0.036)	(0.089)	(0.072)	(0.028)	(0.048)	(0.036)	(0.052)
	N	2220	2121	544	829	2966	1104	1794	1443

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.

Table 4.10: Multimedia Production Function - Period 1

	PC		TV	
	OLSb	IV	OLSb	IV
Peabody Pict. Vocabulary	0.022**	0.053**	-0.002	-0.056**
	(0.005)	(0.012)	(0.003)	(0.018)
Who am I?	0.029**	0.044**	-0.008**	-0.045**
	(0.005)	(0.011)	(0.002)	(0.016)
SDQ Prosocial	0.015**	0.030**	-0.008**	0.006
	(0.005)	(0.012)	(0.003)	(0.017)
SDQ Hyperactivity	0.010	0.018	-0.014**	-0.032
	(0.005)	(0.011)	(0.002)	(0.017)
SDQ Emotional symptoms	-0.000	0.007	-0.012**	-0.032
	(0.005)	(0.011)	(0.003)	(0.017)
SDQ Peers	0.005	0.030**	-0.011**	-0.004
	(0.005)	(0.011)	(0.003)	(0.017)
SDQ Conduct	0.013*	0.038**	-0.016**	-0.015
	(0.005)	(0.011)	(0.003)	(0.017)

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.
IV: instrument computer time with computer access and TV time with number of televisions in the house.

Table 4.11: Multimedia Production Function - Period 2

	PC			TV			VG		
	OLSb	IV	VA	OLSb	IV	VA	OLSb	IV	VA
Peabody Pict. Vocabulary	0.008	0.015	0.007	0.001	-0.064	0.002	-0.018**	-0.016	-0.012**
	(0.004)	(0.016)	(0.004)	(0.002)	(0.050)	(0.002)	(0.004)	(0.021)	(0.004)
Matrix Reasoning	0.013**	0.049**	0.009*	-0.005*	-0.090	-0.002	-0.004	0.036	-0.004
	(0.004)	(0.018)	(0.004)	(0.002)	(0.056)	(0.002)	(0.005)	(0.024)	(0.005)
SDQ Prosocial	-0.006	0.008	-0.004	-0.002	0.023	0.001	0.005	0.008	0.006
	(0.004)	(0.015)	(0.004)	(0.002)	(0.050)	(0.002)	(0.005)	(0.020)	(0.004)
SDQ Hyperactivity	-0.003	0.013	0.000	-0.006**	-0.112	-0.001	-0.000	0.036	-0.002
	(0.004)	(0.018)	(0.003)	(0.002)	(0.058)	(0.002)	(0.005)	(0.024)	(0.004)
SDQ Emotional symptoms	-0.000	0.020	0.001	-0.005*	-0.098	-0.005*	-0.001	0.035	-0.001
	(0.004)	(0.018)	(0.004)	(0.002)	(0.058)	(0.002)	(0.005)	(0.024)	(0.004)
SDQ Peers	-0.015**	-0.004	-0.011**	-0.002	-0.110	-0.000	-0.006	0.061*	-0.008
	(0.004)	(0.019)	(0.004)	(0.002)	(0.060)	(0.002)	(0.005)	(0.025)	(0.004)
SDQ Conduct	-0.005	0.040*	-0.002	-0.008**	-0.130*	-0.004*	-0.002	0.035	0.002
	(0.004)	(0.020)	(0.004)	(0.002)	(0.063)	(0.002)	(0.005)	(0.026)	(0.004)

Standard Errors in brackets. Stars indicate significance at 1% (**) and 5% (*) level.

IV: instrument computer time with computer access, TV time with television in child's bedroom (binary) and video games time with video game console access.

Appendix

Variation in Computer Access and Use Over Time

Table 4.12: Home Computer Access and Use Over Time

	Wave 2						
	No access	No use	<1 hr	1-3 hrs	3-5 hrs	5+ hrs	Total
Wave 1			Weekday				
No access	7.65	7.54	5.18	2.09	0.02	0.02	22.50
No use	1.68	9.15	5.95	1.93	0.02	0.02	18.76
<1 hr	1.97	17.52	22.14	6.89	0.07	0.07	48.67
1-3 hrs	0.38	2.27	3.97	2.98	0.07	0.00	9.67
3-5 hrs	0.00	0.09	0.13	0.09	0.00	0.00	0.31
5+ hrs	0.00	0.00	0.02	0.07	0.00	0.00	0.09
Total	11.69	36.57	37.40	14.05	0.18	0.11	100.00
			Weekend				
No access	7.65	6.30	3.97	4.38	0.18	0.02	22.50
No use	1.79	8.73	6.62	5.81	0.25	0.02	23.22
¡1 hr	1.79	9.94	13.73	14.63	0.43	0.07	40.59
1-3 hrs	0.43	2.22	2.63	6.82	0.81	0.09	12.99
3-5 hrs	0.00	0.11	0.09	0.31	0.09	0.02	0.63
5+ hrs	0.00	0.02	0.02	0.02	0.00	0.00	0.07
Total	11.67	27.33	27.06	31.97	1.75	0.22	100.00

Numbers in table are percentages.

Information used to construct Non-Cognitive skills

Each skill score is equal to the mean of 5 parent-rated items. Some item scores are re-ordered for consistency. Whenever a question changed between the two waves, this is indicated by specifying the wave to which the question refers.

- **SDQ Prosocial**
  1. Considerate of other peoples feelings;
  2. Shares readily with other children (treats, toys, pencils, etc);
  3. Helpful if someone is hurt, upset or feeling ill;
  4. Kind to younger children;
  5. Often volunteers to help others (parents, teachers, other children);
- **SDQ Hyperactivity**
  1. Restless, overactive, cannot stay still for long;
  2. Constantly fidgeting or squirming;
  3. Easily distracted, concentration wanders;
  4. Thinks things out before acting;
  5. Good attention span, sees chores or homework through to the end;
- **SDQ Emotional symptoms**
  1. Often complains of headaches, stomach aches or sickness;
  2. Many worries, often seems worried;
  3. Often unhappy, depressed or tearful;
  4. Nervous or clingy in new situations, easily loses confidence;
  5. Many fears, easily scared;
- **SDQ Peers**
  1. Rather solitary, tends to play alone;
  2. Has at least one good friend;
  3. Generally liked by other children;
  4. Picked on or bullied by other children;
  5. Gets on better with adults than with other children;
- **SDQ Conduct**
  1. Often loses temper;
  2. Generally well behaved, usually does what adults request;
  3. Often fights with other children or bullies them;
  4. Often argumentative with adults (wave 1); Often lies or cheats (wave 2);
  5. Can be spiteful to others (wave 1); Steals from home, school or elsewhere (wave 2);

Control variables used in tables 4.5 and 4.6.

We use the abbreviations 's.c.' (study child), and 'no.' (number).

- **OLSa**:

  FI : family member home activities with the s.c. in the last week (read to s.c. from a book; told s.c. a story not from a book; drawn pictures or did other art or craft activities with s.c.; played music, sang songs, danced or did other musical activities with s.c.; played with toys or games indoors, like board or card games with child; involved child in everyday activities at home, such as cooking or caring for pets; played a game outdoors or exercised together like walking, swimming, cycling); family member outdoor activities with the s.c. in the last month (gone to a movie; gone to a playground or a swimming pool; gone to sporting event in which child was not a player; gone to a live performance for children, like a concert or play; attended a school, cultural or community event; attended a religious service, church, temple, synagogue or mosque; visited a library); s.c. regularly spoken to in a language other than English by parents, babysitters or at child care/pre-school/ school; s.c. regularly attended special or extra cost activities that are not part of his/her normal child care, pre-school or school activities in the last 6 months? (swimming; gymnastics/kindergym; team sport; musical instruments or singing; ballet or other dance; children's religious group; other);

  SI : type of school attended by the s.c. (adjusted by age); grade or year level in school; does child go to a school, kindergarten, pre-school or a day care centre? (wave 1); no. of hours on average per week s.c. goes to (school/ kindergarten/ pre-school/ day care) (wave 1); no. of different schools attended since beginning fulltime schooling (wave 2); computer in school (room has use of a computer; how often do the children have access to the computer).

  OM : TV hours; video game hours (wave 2).

- **OLSb**: OLSa controls +age (child); state of residence; age (parents); s.c. relationship to parents (biological or not); no. of people in home; grandmother in home; grandfather in home; no. of siblings; no. of young siblings; no. of same age siblings; no. of brothers; no. of sisters; no. of younger brother; no. of younger sisters; s.c. has a step-or half-sibling in home; s.c. has an adopted sibling in home; s.c. has a foster sibling in home; parental education; parental work status; financial problems in the last 12 months (could not pay gas, electricity or telephone bills on time; could not pay the mortgage or rent payments on time; went without meals; were unable to heat or cool your home; pawned or sold something because needed cash; sought assistance from a welfare or community organization); parents' annual income; language parents first spoke as a child; country grandparents were born.

Instrumental Variable

Table 4.13: Comparing households with and without a home computer

	Wave 1	Wave 2
Child's Age (months)	-0.024	0.003
Number of Siblings	0.028	-0.091
Father Age (years)	0.713**	0.782*
Mother Age (years)	1.819**	1.803**
Father Higher Education	0.085**	0.100**
Mother Higher Education	0.141**	0.139**
Father Income (10 thous)	1.120**	1.492**
Mother Income (10 thous)	0.211**	0.460**
Mother Employed Full-Time	0.020	0.083**
Mother Employed Part-Time	0.118**	0.121**

Numbers in table are $E(X_t|HC_t = 1) - E(X_t|HC_t = 0)$.
Stars indicate significance at 1% (**) and 5% (*) level.
$HC_1 = 1$: household with home computer.
$HC_1 = 0$: household without home computer.

Table 4.14: First Stage regression

	C_1	C_2
Computer Access (CA)	3.240**	3.474**
	(0.099)	(0.176)
Number of Siblings	-0.041	-0.029
	(0.134)	(0.311)
Number of Younger Siblings	-0.191	-0.212
	(0.199)	(0.356)
Father's Income	0.008	0.024
	(0.013)	(0.017)
Mother's Income	-0.011	-0.013
	(0.025)	(0.029)

Standard Errors in brackets. Star at 1% (**) and 5% (*) level.

BIBLIOGRAPHY

A. Agresti. *Categorical Data Analysis*. John Wiley and Sons, Inc., 2002.

J. Aitchison and S. D. Silvey. Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, 29:813–828, 1958.

J. Angrist and V. Lavy. New evidence on classroom computers and pupil learning. *Economic Journal*, 112(482):735–765, 2002.

J. D. Angrist and G. W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995.

P. Arcidiacono. Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524, 2005.

P. Arcidiacono and J. B. Jones. Finite mixture distributions, sequential likelihood, and the em algorithm. *Econometrica*, 71(2):933–946, 2003.

O. Attanasio, E. Fitzsimons, and C. Meghir. The impact of a conditional education subsidy on the activities of young adults in the uk, 2004. mimeo.

A. Banerjee, S. Cole, E. Duflo, and L. Linden. Remedying education: Evidence from two randomized experiments in india. *Quarterly Journal of Economics*, 122(3):1235–1264, 2007.

F. Bartolucci, R. Colombi, and A. Forcina. An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statistica Sinica*, 17(2), 2007.

F. Bartolucci, A. Forcina, and V. Dardanoni. Positive quadrant dependence and marginal modelling in two-way tables with ordered margins. *Journal of the American Statistical Association*, 96(456):1497–1505, 2001.

G. S. Becker and N. Tomes. An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, 87(6):1153–1189, 1979.

G. S. Becker and N. Tomes. Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(s3):S1–S39, 1986.

P. Belley and L. Lochner. The changing role of family income and ability in determining educational achievement. *Journal of Human Capital*, 1(1):37–90, 2007.

D.O. Beltran, K. K. Das, and R. W. Fairlie. Are computers good for children? the effects of home computers on educational outcomes. Discussion Paper 1912, IZA, 2006.

C. Belzil and J. Hansen. Unobserved ability and the return to schooling. *Econometrica*, 70(2):2078–2091, 2002.

W. P. Bergsma and T. Rudas. Marginal models for categorical data. *Annals of Statistics*, 30(1):140–159, 2002.

S. Berlinski, S. Galiani, and M. Manacorda. Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, 92(5-6):1416–1440, 2008.

R. Bernal and M. P. Keane. Quasi-structural estimation of a model of child care choices and child cognitive ability production. *Journal of Econometrics*, forthcoming, 2008.

A. Bjorklund, M. Lindahl, and E. Plug. The origins of intergenerational associations: Lessons from swedish adoption data. *Quarterly Journal of Economics*, 121(3):999–1028, 2006.

J. Blanden, P. Gregg, and L. Macmillan. Accounting for intergenerational income persistence: Noncognitive skills, ability and education. *Economic Journal*, 117(519):C43–C60, 2007.

F. Bourguignon, F. H. G. Ferreira, and M. Menendez. Inequality of outcomes and inequality of opportunities in brasil. DELTA Working Papers 2003-24, DELTA, 2003.

S. Bowles and H. Gintis. The inheritance of inequality. *Journal of Economic Perspectives*, 16(3):3–30, 2002.

S. V. Cameron and J. J. Heckman. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106(2):262–333, 1998.

S. V. Cameron and J. J. Heckman. The dynamics of educational attainment for black, hispanic and white males. *Journal of Political Economy*, 109(3):455–499, 2001.

S. V. Cameron and C. Taber. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy*, 112(1):132–182, 2004.

P. Carneiro and J. J. Heckman. The evidence on credit constraints in post-secondary schooling. *Economic Journal*, 112 (482):705–734, 2002.

R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement Error in Nonlinear Models*. Chapman Hall, 2006.

R. Colombi and A. Forcina. Marginal regression models for the analysis of positive association. *Biometrika*, 88(4): 1007–1019, 2001.

J. Conlisk. Monotone mobility matrices. *Journal of Mathematical Sociology*, 15:173–191, 1990.

G. Conlon. The differential in earnings premia between academically and vocationally trained males in the united kingdom, 2001. Centre for the Economics of Education Discussion Papers.

F. Cunha and J. J. Heckman. The technology of skill formation. *American Economic Review*, 97(2), 2007.

F. Cunha, J. J. Heckman, L. Lochner, and D. V. Masterov. *Handbook of the Economics of Education*, volume 1, chapter 12, pages 697–805. Elsevier, 2006.

G. B. Dahl and L. Lochner. The impact of family income on child achievement. Working Paper 11279, National Bureau of Economic Research, April 2005.

V. Dardanoni. Measuring social mobility. *Journal of Economic Theory*, 61(2):372–394, 1993.

V. Dardanoni. Income distribution dynamics: Monotone markov chains make light work. *Social Choice and Welfare*, 12(2):181–192, 1995.

V. Dardanoni, G. S. Fields, J. E. Roemer, and M. L. Sanchez Puerta. *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, chapter 3. Stamford University Press, 2006.

V. Dardanoni and A. Forcina. A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association*, 93(443):1112–1123, 1998.

V. Dardanoni and A. Forcina. Inference for lorenz curve orderings. *Econometrics Journal*, 2(1):48–74, 1999.

V. Dardanoni and A. Forcina. Multivariate ordered regression. (mimeo), 2008.

L. Dearden, S. Machin, and H. Reed. Intergenerational mobility in britain. *Economic Journal*, 107(440):47–66, 1997.

L. Dearden, L. McGranahan, and B. Sianesi. The role of credit constraints in educational choices: Evidence from the ncds and bcs70. Working Paper 48, Centre for the Economics of Education, 2004. London School of Economics, London.

J. E. DiNardo and J. P. Pischke. The returns to computer use revisited: Have pencils changed the wage structure too? *Quarterly Journal of Economics*, 112(1):291–303, 1997.

R. Douglas, S. E. Fienberg, M. T. Lee, A. R. Sampson, and L. R. Whitaker. *Topics in Statistical Dependence*, chapter Positive Dependence Concepts for Ordinal Contingency Tables, pages 189–202. Hayward, CA: Institute of Mathematical Statistics, 1990.

R. Dykstra, S. Kochar, and T. Robertson. Statistical inference for uniform stochastic ordering in several populations. *The Annals of Statistics*, 19(2):870–888, 1991.

Z. Eckstein and K. I. Wolpin. The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, 24(4):562–598, 1989.

R. Erikson and J. H. Goldthorpe. Intergenerational inequality: A sociological perspective. *Journal of Economic Perspectives*, 16(3):31–44, 2002.

B. S. Everitt and D. J. Hand. *Finite Mixture Distributions*. Chapman and Hall, London and New York, 1981.

M. Fleurbaey. *Fairness, Responsibility, and Welfare*. Oxford University Press, 2008.

H. B. Ganzeboom, R. Luijkx, and D. J. Treiman. Intergenerational class mobility in comparative perspective. *Research in Social Stratification and Mobility*, 8:3–79, 1989. URL `http://home.fsw.vu.nl/HBG.Ganzeboom/Pdf/Ganzeboom_Luijkx_Treiman_%5BRSSM1989%5D.pdf`.

M. Gentzkow and J. M. Shapiro. Preschool television viewing and adolescent test scores historical evidence from the coleman study. *Quarterly Journal of Economics*, 123(1):forthcoming, 2008.

N. D. Grawe and C. B. Mulligan. Economic interpretations of intergenerational correlations. *Journal of Economic Perspectives*, 16(3):45–58, 2002.

M. Gurian. *Boys and Girls Learn Differently!* John Wiley and Son, 2002.

S. Han and C. B. Mulligan. Human capital, heterogeneity and estimated degrees of intergenerational mobility. *Economic Journal*, 111(470):207–43, 2001.

J. J. Heckman, L. Lochner, and C. Taber. General-equilibrium treatment effects: A study of tuition policy. *American Economic Review*, 88(2):381–386, 1998.

J. J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320, 1984.

J. J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 2006.

R. J. Herrnstein and C. Murray. *The Bell Curve*. The Free Press, 1994.

H. Ichimura and C. Taber. Semiparametric reduced-form estimation of tuition subsidies. *American Economic Review*, 92(2):286–292, 2002.

M. P. Keane and K. I. Wolpin. The career decisions of young men. *Journal of Political Economy*, 105(3):473–522, 1997.

M. P. Keane and K. I. Wolpin. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103, 2001.

J. Keilson and A. Kester. Monotone matrices and monotone markov chains. *Stochastic Processes and their Applications*, 5:231–41, 1977.

A. B. Krueger. How computers have changed the wage structure: Evidence from microdata, 1984-1989. *Quarterly Journal of Economics*, 108(1):33–60, 1993.

Donghoon Lee. An estimable dynamic general equilibrium model of work, schooling and occupational choice. *International Economic Review*, 46(1):1–34, 2005.

A. Lefranc, N. Pistolesi, and A. Trannoy. Inequality of opportunities vs. inequality of outcomes: Are western societies all alike? Working Papers 54, ECINEQ, Society for the Study of Economic Inequality, 2006.

E. Lehmann. Ordered families of distributions. *Annals of Mathematical Statistics*, 26:399–419, 1955.

T. Magnac and D. Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002.

G. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley and Sons, INC., 2000.

C. B. Mulligan. Galton versus the human capital approach to inheritance. *Journal of Political Economy*, 107(6): S184–S224, 1999.

J. M. Neuhaus. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86(4):843–855, 1999.

V. Peragine. Ranking income distributions according to equality of opportunity. *Journal of Economic Inequality*, 2(1): 11–30, April 2004.

S. J. Prais. Measuring social mobility. *Journal of the Royal Statistical Society: Series A*, 118:56–66, 1955.

L. Putterman, J. E. Roemer, and J. Silvestre. Does egalitarianism have a future? *Journal of Economic Literature*, 36 (2):861–902, June 1998.

D. Restuccia and C. Urrutia. Intergenerational persistence of earnings: The role of early and college education. *American Economic Review*, 94(5):1354–78, 2004.

S. G. Rivkin, E. A. Hanushek, and J. F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2): 417–458, 2005.

T. Robertson and F. T. Wright. Bounds on mixtures of distributions arising in order restricted inference. *The Annals of Statistics*, 10(1):302–306, 1982.

T. Robertson, F.T. Wright, and R. Dykstra. *Order Restricted Statistical Inference*. J. Wiley and Sons, New York, N.Y., 1988.

J. E. Roemer. *Equality of Opprotunity*. Meritocracy and Economic Inequality. Princeton University Press, 2000. editors K. Arrow, Samuel Bowles and S. Durlauf.

C. E. Rouse, A. B. Krueger, and L. Markman. Putting computerized instruction to the test: a randomized evaluation of a scientifically-based reading program. *Economics of Education Review*, 23(4):323–338, 2004.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economics Papers*, 3:135–146, 1951.

J. Rust. *Structural Estimation of Markow Decision Processes*, volume IV of *Handbook of Econometrics*, chapter 51. Elsevier Science, 1994. editors R.F. Engle and D.L. McFaddden.

M. Evans Schmidt and D. R. Anderson. *Children And Television: Fifty Years of Research*, chapter 3, pages 65–84. Routledge, 2007.

J. Schmitt and J. Wadsworth. Is there an impact of household computer ownership on childrens educational attainment in britain? *Economics of Education Review*, 25(6):659–673, 2006.

A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.

A. Shapiro. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72(1):133–144, 1985.

M. J. Silvapulle and P. K. Sen. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. 2005.

G. Solon. Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 29, pages 1761–1800. Elsevier, October 1999.

K. Subrahmanyam, P. Greenfield, R. Kraut, and E. F. Gross. The impact of computer use on children's and adolescents' development. *Journal of Applied Developmental Psychology*, 22(1):7–30, 2001.

K. Subrahmanyam, R. E. Kraut, P. M. Greenfield, and E. F. Gross. The impact of home computer use on children's activities and development. *The Future of Children*, 10(2), 2000.

P. E. Todd and K. I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113(485):3–33, 2003.

P. E. Todd and K. I. Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, 1(1):91–136, 2007.

R. J. Willis and S. Rosen. Education and self-selection. *Journal of Political Economy*, 87(5):S7–S36, 1979.

F. A. Wolak. The local nature of hypotheses tests involving inequality constraints,. *Econometrica*, 59(4):981–995, 1991.

D. J. Zimmerman. Regression toward mediocrity in economic stature. *American Economic Review*, 82(3):409–429, 1992.